

WO9849274

Publication Title:

THERMOSTABLE DNA POLYMERASE AND INTEINS OF THE
THERMOCOCCUS FUMICOLANS SPECIES

Abstract:

Abstract of WO9849274

The invention concerns a purified thermostable DNA polymerase, thermostable archae bacteria DNA polymerase of the Thermococcus fumicolans species having a molecular weight of the order of 89000 daltons and its thermostable inteins. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

ATTORNEY DOCKET NUMBER: 10589-034-999
SERIAL NUMBER: 10/551,300
REFERENCE: **B06**

PCT

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
Bureau international



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

| | | |
|---|-----------|--|
| (51) Classification internationale des brevets ⁶ : C12N 9/12 // 15/54 | A1 | (11) Numéro de publication internationale: WO 98/49274 (43) Date de publication internationale: 5 novembre 1998 (05.11.98) |
| <p>(21) Numéro de la demande internationale: PCT/FR97/00761</p> <p>(22) Date de dépôt international: 29 avril 1997 (29.04.97)</p> <p>(71) Déposant (pour tous les Etats désignés sauf US): APPLI-GENE-ONCOR [FR/FR]; Parc d'Innovation, Rue Geiler de Kaysersberg, Boîte postale 72, F-67402 Illkirch (FR).</p> <p>(72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): QUERELLOU, Joël [FR/FR]; L'Arc'Hantel, F-29280 Brest (FR). CAMBON, Marie-Anne [FR/FR]; 8, allée de Stelle, F-29280 Plouzane (FR).</p> <p>(74) Mandataire: BREESE-MAJEROWICZ; 3, avenue de l'Opéra, F-75001 Paris (FR).</p> | | <p>(81) Etats désignés: CA, JP, US, brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Publiée <i>Avec rapport de recherche internationale.</i></p> |
| <p>(54) Title: THERMOSTABLE DNA POLYMERASE AND INTEINS OF THE THERMOCOCCUS FUMICOLANS SPECIES</p> <p>(54) Titre: ADN POLYMERASE THERMOSTABLE ET INTEINES DE L'ESPECE THERMOCOCCUS FUMICOLANS</p> <p>(57) Abstract</p> <p>The invention concerns a purified thermostable DNA polymerase, thermostable archae bacteria DNA polymerase of the <i>Thermococcus fumicolans</i> species having a molecular weight of the order of 89000 daltons and its thermostable inteins.</p> <p>(57) Abrégé</p> <p>La présente invention concerne une ADN polymérase purifiée thermostable ADN polymérase thermostable d'archaebactéries de l'espèce <i>Thermococcus fumicolans</i> ayant un poids moléculaire de l'ordre de 89 000 daltons, ainsi que ses intéines thermostables.</p> | | |

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

| | | | | | | | |
|----|---------------------------|----|---|----|--|----|-----------------------|
| AL | Albanie | ES | Espagne | LS | Lesotho | SI | Slovénie |
| AM | Arménie | FI | Finlande | LT | Lituanie | SK | Slovaquie |
| AT | Autriche | FR | France | LU | Luxembourg | SN | Sénégal |
| AU | Australie | GA | Gabon | LV | Lettonie | SZ | Swaziland |
| AZ | Azerbaïdjan | GB | Royaume-Uni | MC | Monaco | TD | Tchad |
| BA | Bosnie-Herzégovine | GE | Géorgie | MD | République de Moldova | TG | Togo |
| BB | Barbade | GH | Ghana | MG | Madagascar | TJ | Tadjikistan |
| BE | Belgique | GN | Guinée | MK | Ex-République yougoslave de Macédoine | TM | Turkménistan |
| BF | Burkina Faso | GR | Grèce | | | TR | Turquie |
| BG | Bulgarie | HU | Hongrie | ML | Mali | TT | Trinité-et-Tobago |
| BJ | Bénin | IE | Irlande | MN | Mongolie | UA | Ukraine |
| BR | Brésil | IL | Israël | MR | Mauritanie | UG | Ouganda |
| BY | Bélarus | IS | Islande | MW | Malawi | US | Etats-Unis d'Amérique |
| CA | Canada | IT | Italie | MX | Mexique | UZ | Ouzbékistan |
| CF | République centrafricaine | JP | Japon | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Pays-Bas | YU | Yougoslavie |
| CH | Suisse | KG | Kirghizistan | NO | Norvège | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | République populaire démocratique de Corée | NZ | Nouvelle-Zélande | | |
| CM | Cameroun | | | PL | Pologne | | |
| CN | Chine | KR | République de Corée | PT | Portugal | | |
| CU | Cuba | KZ | Kazakhstan | RO | Roumanie | | |
| CZ | République tchèque | LC | Sainte-Lucie | RU | Fédération de Russie | | |
| DE | Allemagne | LI | Liechtenstein | SD | Soudan | | |
| DK | Danemark | LK | Sri Lanka | SE | Suède | | |
| EE | Estonie | LR | Libéria | SG | Singapour | | |

ADN POLYMERASE THERMOSTABLE ET INTEINES DE L'ESPECE
THERMOCOCCUS FUMICOLANS

La présente invention concerne une nouvelle ADN
5 polymérase thermostable ainsi que ses deux intéines,
provenant d'une archaebactérie de l'espèce *Thermococcus*
fumicolans.

Les ADN polymérases sont des enzymes impliquées
dans la répllication et la réparation de l'ADN dans toute
10 cellule vivante. On connaît aujourd'hui de nombreuses ADN
polymérases isolées de micro-organismes tel que *E. coli*
(ADN polymérase I) ou du phage T4. Des ADN polymérases
ont aussi été identifiées et purifiées et à partir de
micro-organismes thermophiles comme *Thermus aquaticus*
15 (Taq polymérase, Chien, A. et al. J. Bact. 1976,
127:1550-1557 ; Kaladin et al. Biokhymiya 1980, 45:644-
651), *Thermus thermophilus*, ou encore des espèces du
genre *Bacillus* (demande de brevet Européen publiée sous
le No. 699 760), *Thermococcus* (demande de brevet Européen
20 No. 455 430), *Sulfolobus* et *Pyrococcus* (demande de brevet
Européen publiée sous le No. 547 359). Parmi ces ADN
polymérases issues d'archaebactéries on peut citer la
Pfu, isolées de *Pyrococcus furiosus* (18), la Vent™
polymerase de *Thermococcus litoralis* (10), la 9°N de
25 *Pyrococcus sp. 9°N* (15) et la DeepVent™ de *Pyrococcus GB-*
D, les deux premières provenant de souches du littoral
(Baie de Naples), les deux suivantes de souches sous-
marines profondes.

Le mécanisme d'action des ADN polymérases est
30 aujourd'hui relativement bien connu et consiste en une
réplication de l'ADN à l'identique selon un mode semi-
conservatif. Le brin recopié sert de matrice et les
quatre nucléotides triphosphates sont le substrat de
cette polymérisation. Les enzymes ayant une activité ADN
35 polymérase sont aujourd'hui de plus en plus utilisées *in*
vitro afin de travailler en biologie moléculaire dans
divers buts tels le clonage, la détection d'erreurs, le

séquençage, le marquage, et de façon générale, l'amplification de séquences d'acides nucléiques.

5 Cette amplification, *in vitro*, de séquences d'acide désoxyribonucléique fait appel à la technique de la réaction de polymérisation en chaîne (PCR) décrite dans les brevets Européens No. 200 362 et 201 184. Le principe de cette technique est basé sur la réalisation de cycles successifs d'extension d'amorces mettant en oeuvre les quatre nucléotides triphosphates ainsi qu'une
10 ADN polymérase et un ADN matrice à recopier. A chaque cycle, l'enzyme double le nombre de brins d'ADN disponibles et entre chaque cycle une thermodénaturation est nécessaire afin d'ouvrir la double hélice d'ADN pour le cycle suivant. Les températures utilisées pour cette
15 étape de thermodénaturation ne sont pas compatibles avec la conservation de l'activité de la plupart des ADN polymérases connues, telle la Klenow. C'est ainsi que des nombreux efforts de recherche ont été réalisés afin de trouver des enzymes supportant ces températures.

20 Cependant, si les micro-organismes thermophiles sont aujourd'hui connus, il reste encore difficile d'obtenir ces enzymes thermostables avec des rendements de production suffisants. La biologie moléculaire et le génie génétique permettent de palier cet inconvénient.
25 Ainsi, une fois repéré dans le génome, le gène codant pour l'ADN polymérase est cloné, séquencé puis reclone dans un vecteur d'expression afin de produire la protéine dite alors recombinante, chez un hôte mésophile plus aisé à cultiver tel *E. coli* ou *S. cerevisiae*. Cette méthode
30 d'expression chez *E. coli* a notamment été décrite dans la demande de brevet internationale PCT publiée sous le No. WO 89/06691 pour produire l'ADN polymerase de *Thermus aquaticus*.

35 L'ADN polymérase de l'invention provient d'une archaebactérie de l'espèce *Thermococcus fumicolans*. Outre ses propriétés thermostables la rendant particulièrement efficace notamment dans une processus de PCR, cette ADN

polymérase est remarquable en ce qu'elle présente deux "introns protéiques", encore appelés "intéines", au niveau de son polypeptide précurseur.

La séquence nucléotidique de ses intéines est
5 insérée dans celle de l'ADN polymérase, généralement au niveau de sites conservés impliqués, après traduction, dans les réactions catalytiques. Ces séquences sont transcrites et traduites en même temps que celle de l'ADN polymérase et l'épissage autocatalytique des intéines
10 produit alors trois enzymes: deux intéines et une ADN polymérase.

On trouve de telles intéines également au sein d'autres molécules telles l'ATPase vacuolaire chez *S. cerevisiae* (4), GyrA chez *Mycobacterium leprae* (7), Rec A
15 chez *Mycobacterium tuberculosis* (5, 6). Les intéines font partie pour leur majorité de la famille des endonucléases de type "homing endonucleases" puisqu'elles coupent l'ADN en un site reconnu, à l'endroit même où leur séquence nucléotidique vient s'insérer.

20 Le développement des biotechnologies tant dans la recherche que dans les domaines de la médecine ou de l'agro-alimentaire, nécessite de disposer de divers types d'ADN polymérases susceptibles d'améliorer quantitativement et qualitativement des techniques aussi
25 diverses que le clonage, la détection, l'amplification de séquences d'ADN. La présente invention vise précisément à offrir une nouvelle ADN polymérase thermostable qui est issue d'une espèce récemment décrite: *Thermococcus fumicolans* (8). Cet isolat a été isolé à partir de
30 fragments de cheminées prélevées dans le bassin Nord-Fidgien lors de la campagne franco-japonnaise STARMER en 1989. Cette espèce, anaérobie stricte, présente une température optimale de croissance de 90°C, ce qui est relativement élevé pour un *Thermococcus*. Son pH optimum
35 est de 8,8, et son taux de salinité de 20 g/l à 40 g/l.

L'invention a donc pour objet une ADN polymérase purifiée thermostable d'archaebacteries de

l'espèce *Thermococcus fumicolans* ayant un poids moléculaire de l'ordre de 89000 Da ainsi que ses intéines thermostables, dont le gène comportant les deux séquences codant pour lesdites intéines a été cloné.

5 Les travaux de recherche ayant permis d'identifier, de séquencer et d'étudier l'ADN polymérase de l'invention ont été réalisés à partir de la souche *Thermococcus fumicolans* ST557 déposée à la Collection de l'Institut Pasteur (CIP) sous le numéro CIP 104680. Cette
10 ADN polymérase sera dénommée dans ce qui suit *Tfu*. Sa séquence de 774 acides aminés est représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2. Un poids moléculaire de 89797 Da et un pI de 8.1 ont été déduits de cette séquence.

15 L'invention concerne donc l'ADN polymérase purifiée thermostable d'archaebactéries de l'espèce *Thermococcus fumicolans* ayant un poids moléculaire de l'ordre de 89 000 daltons ainsi que ses dérivés
20 enzymatiquement équivalents. On entend par dérivés enzymatiquement équivalent, les polypeptides et protéines constitués par ou comprenant la séquence en acides aminés représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2 dès lors qu'ils présentent les
25 propriétés de l'ADN polymérase de *Thermococcus fumicolans*. A ce titre l'invention envisage plus particulièrement une ADN polymérase dont la séquence en acides aminés est représentée dans la liste de séquence en annexe sous le numéro SEQ ID NO:1 ou un fragment de
30 celle-ci ou encore un assemblage de tels fragments, comme la séquence de 774 acides aminés représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:2.

En effet, la présence de deux intéines I-Tfu-1 et I-Tfu-2 dans la séquence numéro SEQ ID NO:1, sont
35 susceptibles de conduire lors de la préparation par synthèse chimique ou par génie génétique, à des séquences d'ADN polymérase de *T. fumiculans* tronquées dont les

propriétés enzymatiques sont équivalente à celle de l'ADN polymérase de *T. fumicolans* purifiée.

On entend aussi par dérivés, les séquences en acides aminés ci-dessus modifiées par insertion et/ou
5 déléation et/ou substitution d'un ou plusieurs aminoacides, pour autant que les propriétés de l'ADN polymérase de *T. fumicolans* qui en résultent ne soient pas significativement modifiées.

L'invention concerne également une séquence
10 d'ADN constituée par ou comprenant la séquence codant pour une ADN polymérase de l'invention.

La séquence d'ADN représenté dans la liste de séquences en annexe sous le numéro SED ID NO: 1 représente une telle séquence. L'ADN codant pour l'ADN
15 polymérase de *T. fumicolans* et ses deux intéines est constituée par le nucléotides 357 à 5028. Les nucléotides 1 à 356 correspondent à la région promotrice de ce gène. En conséquence, l'invention a pour objet une séquence d'ADN constituée par ou comprenant la séquence comprise
20 entre les nucléotides 357 à 5028 de la SED ID NO: 1, ou un fragment de celle-ci ou encore un assemblage de tels fragments.

l'invention se rapporte tout particulièrement à une séquence d'ADN constituée par ou comprenant les
25 nucléotides 357 à 1674 et 2755 à 3156 et 4324 à 5028 de la séquence d'ADN représenté dans la liste de séquences en annexe sous le numéro SED ID NO: 1.

Cette séquence code pour l'ADN polymérase de *T. fumicolans* dont la séquence de 774 acides aminés est
30 représentée dans la liste de séquences en annexe sous le numéro SED ID NO: 2.

L'invention concerne autant l'ADN polymérase isolées et purifiées de la souche *Thermococcus fumicolans* que l'ADN polymérase préparées par synthèse chimique, par
35 exemple par ligature de fragments polypeptidiques, ou encore par les méthodes du génie génétique.

Dans le cadre de ces méthodes du génie génétique, l'invention a aussi pour objet un vecteur comprenant une séquence d'ADN définie précédemment, ainsi qu'un procédé de production ou d'expression dans un hôte cellulaire des ADN thermostables de l'invention.

Un procédé de production d'une ADN polymérase conforme à l'invention consiste:

- à transférer une séquence d'acide nucléique codant pour l'ADN polymérase ou un vecteur contenant ladite séquence dans un hôte cellulaire,
- à cultiver l'hôte cellulaire obtenu à l'étape précédente dans des conditions permettant la production de l'ADN polymérase,
- à isoler, par tous moyens appropriés, ladite ADN polymérase.

L'hôte cellulaire mis en oeuvre dans les procédés précédents peut être choisi parmi les procaryotes ou les eucaryotes et notamment parmi les bactéries, les levures, les cellules de mammifères, de plantes ou d'insectes.

Le vecteur utilisé est choisi en fonction de l'hôte dans lequel il sera transféré.

L'ADN polymérase thermostable de l'invention est utile notamment dans les procédés d'amplification enzymatique de séquences d'acides nucléiques. En conséquence, l'invention a pour objet, de tels procédés mettant en oeuvre l'ADN polymérase thermostable décrite précédemment, ainsi que les kits d'amplification comprenant, outre les réactifs généralement utilisés, une quantité adéquate de cette ADN polymérase.

L'invention a également pour objet une intéine purifiée thermostable d'archaebactéries de l'espèce *Thermococcus fumicolans*.

Comme indiqué précédemment, les intéines sont aussi définies comme des introns protéiques qui sont non pas épissés au niveau de l'ARN messager mais au niveau

d'une maturation protéique. Elles relèvent donc d'un seul gène traduit et transcrit en une seule étape, et constituent des sous produits de la maturation de la protéine codée par ce gène (Xu, M.G., Comb, D.G., Paulus H., Noren C.J., Shao Y., Perler, F., 1994, *Protein splicing : an analysis of the branched intermediate and its resolution by succinimidine formation*. *EMBO J* . 13, 5517-5522.)

Les intéines sont des endonucléases de restriction qui ont la propriété de couper l'ADN à l'endroit même où s'insère leur gène, et en conséquence, elles peuvent être considérées comme des séquences égoïstes.

Les intéines possèdent dans leur séquence toute l'information nécessaire à leur propre épissage puisqu'elles s'épissent dans *E. coli*.

Il est possible de distinguer quatre grandes étapes de maturation protéique :

- Une première étape de formation d'un intermédiaire linéaire qui possède une fonction ester. Cette réaction est dépendante du pH et de l'environnement local de cette liaison (nature des acides aminés). Ce principe est utilisé dans les kits de clonage, d'expression ou de purification mettant en oeuvre des intéines, car un changement de l'environnement provoque ou non un épissage. En effet, celui-ci serait inhibé à pH 11 et activé à pH 7,5.

- Une deuxième étape de transestérification qui permet de transformer l'intermédiaire précédent et de déplacer l'équilibre de la première étape.

- La troisième étape consiste en une cyclisation de l'asparagine libérant l'intéine.

- La quatrième étape est la stabilisation de la protéine mature et la formation d'une réelle liaison peptidique.

Il est donc possible de construire des mutants thermosensibles permettant de bloquer l'épissage

protéique aux températures d'expression (30°C) et de l'induire par chauffage.

Cette possibilité de contrôle de l'épissage protéique par la température peut-être utilisée dans des vecteurs de clonage avec une séquence codant pour l'intéine et des sites de clonage autour. Si la protéine à cloner et à exprimer est toxique pour l'hôte, on peut la cloner en deux fragments autour de la séquence d'intéine. Ainsi, globalement, le gène à cloner est complet mais il est interrompu par la séquence de l'intéine. Lors de l'expression, l'intéine se retrouve dans la protéine exprimée, la rendant ainsi inactive. Il est alors possible par chauffage, à la fin de l'induction, de libérer l'intéine par épissage autocatalytique et ainsi retrouver la protéine clonée active.

Les intéines permettent ainsi de réaliser des purifications et sont utilisés dans des kits selon le principe décrit ci-après. Certains résidus autour de site d'épissage de l'intéine sont modifiés. L'expression de la protéine recombinante est réalisée à basse température pour bloquer un éventuel épissage trop précoce. En C-terminal de l'intéine est fixé un site ayant une forte affinité pour la chitine. Lors de l'induction, la protéine clonée est exprimée ainsi que l'intéine et le site de fixation à la chitine. La purification est alors réalisée avec la chitine, sur des colonnes d'affinité, qui retiennent la chitine et aussi l'intéine et la protéine clonée, le tout faisant partie de la même pré-protéine. On hydrolyse alors l'extrémité N-terminale de l'intéine avec du DTT ou du β -mercaptoéthanol pour libérer la protéine clonée.

Les intéines sont aussi des endonucléases de restrictions thermostables, qui ont pour site de reconnaissance l'endroit même où s'insère leur gène dans la séquence "hôte". Elles possèdent une séquence nucléique répétée deux fois (LAGLIDADG) dans la protéine,

séquence plus ou moins conservée et qui correspond au site actif de reconnaissance et de coupure de l'ADN. Ces enzymes semblent aussi avoir besoin de Mg^{++} pour leur activité.

5 Il convient de remarquer que les deux intéines de l'invention sont coexprimées dans *E. coli* et sont autoépissées. Ce qui signifie qu'elle n'ont pas de toxicité pour l'hôte, à la différence de l'une des intéines de *Thermococcus litoralis* (9), et en conséquence
10 leur utilisation dans des kits d'expression ou de purification est aisée.

Une première séquence d'intéine selon l'invention est représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:3. Cette intéine,
15 dénommée I-Tfu-1, présente un poids moléculaire de 41 409 Da et un pI de 9,13, déduits de la séquence de 360 acides aminés de la séquence SEQ ID NO:3.

Une seconde séquence d'intéine selon l'invention est représentée dans la liste de séquences en annexe sous le numéro SEQ ID NO:4. Cette intéine,
20 dénommée I-Tfu-2, présente un poids moléculaire de 44 765 Da et pI de 9,6, déduits de la séquence de 389 acides aminés de la séquence SEQ ID NO:4.

Comme rappelé précédemment, les intéines
25 thermostables de l'invention sont utiles notamment dans les procédés de restriction d'acides nucléiques et dans la mise au point de vecteurs d'expression permettant de réduire la toxicité de la protéine à exprimer en insérant l'une des deux séquences d'intéines dans la séquence de
30 la protéine à exprimer. Ceci peut se faire sans manipulation de la séquence des intéines si le clonage s'effectue chez *E. coli*, les techniques d'expression utilisées ayant démontré leur innocuité pour cet organisme hôte. En conséquence, l'invention a pour objet, de tels
35 procédés mettant en oeuvre l'une des deux ou les deux intéines thermostables décrites précédemment, ainsi que les kits d'expression ou de purification contenant l'une

ou les deux séquences codant pour les dites intéines thermostables.

5 L'invention concerne également une séquence d'ADN constituée par ou comprenant la séquence codant pour une intéine de l'invention.

10 Une séquence d'ADN codant pour l'intéine I-Tfu-1 est comprise entre les nucléotides 1675 et 2754 dans la séquence SED ID NO: 1 en annexe. Cette séquence d'ADN code pour l'intéine dont la séquence en acides aminés est représentée dans la liste de séquences en annexe sous le numéro SED ID NO: 3.

15 Une séquence d'ADN codant pour l'intéine I-Tfu-2 est comprise entre les nucléotides 3157 et 4323 dans la séquence SED ID NO: 1 en annexe. Cette séquence d'ADN code pour l'intéine dont la séquence en acides aminés est représentée dans la liste de séquence en annexe sous le numéro SED ID NO: 4.

20 L'invention concerne autant ces intéines thermostables isolées et purifiées de la souche *Thermococcus fumicolans* que des intéines préparées par synthèse chimique, par exemple par ligature de fragments polypeptidiques, ou encore par les méthodes du génie génétique.

25 Dans le cadre de ces méthodes du génie génétique, l'invention a aussi pour objet un vecteur comprenant une séquence d'ADN définie précédemment, ainsi qu'un procédé de production ou d'expression dans un hôte cellulaire des ADN codant pour les intéines de 30 l'invention. De tels procédés sont identiques à ceux rapportés précédemment pour l'ADN polymérase de *T. fumicolans*.

35 D'autres avantages et caractéristiques de l'invention apparaîtront à la lecture des exemples qui suivent, donnés à titre non limitatifs et concernant le clonage, l'expression, la caractérisation et l'activité

de l'ADN thermostable de l'invention, et se référant aux dessins annexés dans lesquels :

- La figure 1 représente l'hybridation ADN-ADN de l'ADN génomique de *T. fumicolans* digéré par diverses enzymes de restriction et hybridé avec les sondes GE23ClaI-HindIII et GE23XhoI-SalI.

- La figure 2 représente la stratégie de clonage, la structure du gène de l'ADN polymérase de *T. fumiculans* et les produits du gène.

- La figure 3 représente les résultats de purification de la polymérase recombinante de *T. fumicolans* après une colonne d'héparine sépharose, visualisés par SDS-PAGE.

- La figure 4 représente les résultats de purification de la polymérase recombinante de *T. fumiculans* après une colonne Bleue HTrap n°2, visualisé par SDS-PAGE.

- La figure 5 représente les résultats de purification de la polymérase recombinante de *T. fumicolans* après une colonne de phosphocellulose, visualisés par SDS-PAGE.

- La figure 6 représente les résultats de purification de la polymérase recombinante de *T. fumicolans* après une colonne MonoQ, visualisés par SDS-PAGE.

- La figure 7 représente les résultats de PCR avec l'ADN polymérase de *T. fumicolans* avec les fractions d'exclusion de la colonne MonoQ.

I- Matériel et méthodes.

1) Conditions de culture, plasmides et souches utilisées.

Les souches *Thermococcus litoralis* (DSM 5474^T) et *Pyrococcus furiosus* (DSM 3638^T) ont été obtenues auprès de la collection du Deutsche Sammlung von Microorganismen (DSM) Braunschweig-Stocheim, Allemagne.

La souche *Pyrococcus* sp. GE 23 a été isolée de cheminées de sources hydrothermales profondes et a été fournie par G Erauso (CNRS, Station Biologique de Roscoff, France). La souche *Thermococcus fumicolans* a été obtenue auprès du
5 laboratoire de Microbiologie Marine de G. Barbier (IFREMER-DRV-VP-CMM) à Brest, France. Cette souche, *Thermococcus fumicolans* a été obtenue par purification à partir de fragments de cheminées hydrothermales
10 recueillies dans le bassin nord-Fidgien lors de la campagne franco-japonaise STARMER effectuée en 1989 à 2000 mètres de profondeur.

Pyrococcus sp. GE23 a été cultivée à 85°C dans un milieu 2216S (DIFCO) à un pH de 6,5.

Thermococcus fumicolans, décrite en 1996
15 (Godfroy et al.) est cultivée en conditions anaérobies dans un milieu contenant les éléments suivants: peptone 2g/l; extraits de levure 0,5g/l; sel de mer (Sigma) 30g/l; tampon PIPES 6,05g/l, soufre élémentaire 10g/l, rézasurine 1mg/l. Le pH est ajusté à 8,5 par de la soude
20 5N à 20°C.

La souche de *Escherichia coli* SURE (Stratagene, La Jolla, Calif.) a été cultivée dans du milieu LB avec le ou les antibiotiques appropriés, à 37°C sous agitation. Cette souche a été utilisée comme hôte pour
25 recevoir les constructions primaires à partir des vecteurs pUC 18 ou pBluescript. Les souches NovaBlue, BL21(DE3) et BL21(DE3)pLyss (Novagen, Madison, Wi.) ont été cultivées dans du milieu 2xYT avec les antibiotiques appropriés à 37°C ou 30°C. Ces souches ont été utilisées
30 comme hôtes pour les plasmides d'expression.

2) Isolement de l'ADN, hybridation et ADN recombinants.

L'ADN de haut poids moléculaire de *Thermococcus*
35 *fumicolans* a été isolé par la méthode de Charbonnier modifiée (3). Les cellules ont été resuspendues dans un tampon TE-Na 1X, puis lysées à 40°C pendant trois heures

avec un mélange de N-Lauryl-sarcosine 1%, dodecyl sulfate de sodium 1% et 0,4 mg/ml de protéinase K. Après la lyse, une centrifugation à 5000g pendant 10 minutes permet d'éliminer les débris cellulaires. L'ADN est extrait par un traitement au Phenol-Chloroforme-alcool Isoamylique ou PCI (25-24-1), puis traité par la RNase à 5µg/ml à 60°C pendant une heure. Ces étapes sont suivies de deux extractions additionnelles par PCI et d'une extraction au chlorophorme. L'ADN est ensuite précipité à l'éthanol absolu à -20°C, puis centrifugé, séché à l'air et repris dans un tampon TE-1x. La concentration et la pureté de cet ADN sont mesurées par spectrophotométrie à 230, 260 et 280nm avec un appareil GeneQuantII (Pharmacia, Upsalla, Sweden). Pour la construction de la mini-banque génomique en pUC 18 (17), l'ADN a été totalement digéré pendant une nuit à 37°C par une série d'enzymes de restriction (BamHI, HindIII, EcoRI, EcoRV, PvuII, SalI, XbaI et XhoI) par simples et doubles digestions. Puis les fragments d'ADN sont fractionnés sur gel d'agarose à 0,8% dans de TBE-1X et transférés sous vide sur une membrane de nylon Hybond-N+ (Amersham, UK). Ces membranes ont été hybridées avec des sondes d'ADN préparées par PCR avec des amorces spécifiques sélectionnées à partir des gènes d'ADN polymérases de *P. furiosus*, *T. litoralis* et *Pyrococcus* sp. GE23. Ces sondes sont préalablement marquées au ³²P par "random priming" conformément aux recommandations du fabricant (Megaprime, Amersham, UK). Deux sondes de *P. furiosus* ont été utilisées, Pfu et Pfu F, couvrant respectivement les régions délimitées par les paires de bases 8 à 2316 et 819 à 1915 de la section codante du gène de la polymérase comme défini par Uemori et al. (18). Deux sondes de *T. litoralis*, Tli I et Tli T, couvrant respectivement les régions délimitées par les paires de bases 297 à 1768 et 4631 à 5378, comme défini par Hodges et al. (9). Deux sondes de *Pyrococcus* sp GE23 ont été utilisées, l'une contenant la partie 5' du gène (fragment ClaI-HindIII

correspondant aux sites pb 8 et pb 1353 de la section
codante) et l'autre contenant la partie terminale de ce
même gène obtenu par PCR (amorces dites d'expression NdeI
et SalI correspondant aux sites pb -1 et pb 2318, puis
5 digestion par XhoI et SalI comprenant les bases du n°
1879 à 2318). Les fragments positifs ont été identifiés
par hybridation ADN-ADN (14). Seules les hybridations
avec les sondes de *Pyrococcus* sp. GE 23 ont fourni des
signaux positifs à 55°C en moins de 24 h d'exposition.
10 Les sondes issues de *T. litoralis* et de *P. furiosus* n'ont
donné aucun résultat, même à 50°C dans un tampon standard
sans formamide. A partir des hybridations avec les sondes
de *Pyrococcus* sp. GE 23, des fragments HindIII de 1,9 kb
ont été sélectionnés, puis préparés par une digestion
15 appropriée de 100µg d'ADN génomique, purifiés dans des
sacs de dialyse à partir des gels d'agarose, et
précipités à l'éthanol absolu après une extraction au
PCI. Les fragments ont été ligaturés dans un pUC
18/HindIII déphosphorylé. Les transformations des souches
20 hôtes ont été réalisées par électroporation (Gene Pulser,
Biorad). Le criblage des clones recombinants a été
effectué par sélection à l'ampicilline, alpha-
complémentation sur substrat X-Gal-IPTG puis hybridation
de colonies selon les techniques standards (12). La
25 température des hybridations de colonies était de 55°C
avec les sondes de *Pyrococcus* sp GE23, dans un tampon
standard sans formamide. L'ADN plasmidique a été isolé
selon la méthode décrite par Birnboim et Doly (1), puis
purifié par chromatographie échangeuse d'anions en phase
30 solide (Quiagen, Chatsworth, Calif.). Les fragments de
restriction des plasmides ont été purifiés sur gel
d'agarose par la méthode du GeneClean (Bio 101, La Jolla,
Calif.) pour un clonage ultérieur.

L'ARNr 16S et 23S de *Thermococcus fumicolans* a
35 été amplifié par PCR en utilisant les amorces suivantes :
- amorce directe Aa: 5' TCCGGTTGATCCTGCCGGAA-3'
- amorce reverse 23Sa: 5'-CTTTCGGTCGCCCTACT-3'

- étape initiale 3 minutes à 94°C suivi de 30 cycles (94°C, 1min/ 49°C, 1mn/ 72°C, 2mn) et, élongation finale de 5 mn à 72°C.

5 Le produit de PCR a été cloné dans le vecteur pUC18 pour séquençage ultérieur.

3) Séquencage des ADN.

10 Les séquences d'ADN ont été obtenues par la méthode de terminaison de chaîne (13) en utilisant un système d'analyses automatique d'ADN Applied Biosystems. Les deux brins des gènes codant pour l'ADN polymérase et les deux intéines ont été séquencés en utilisant des amorces universelles localisées sur des vecteurs ainsi que des amorces internes.

15 La séquence d'ADNr 16S a été réalisée sur les deux brins, après clonage (SureClone, Pharmacia, Uppsala, Sweeden), en utilisant le kit Hot-Tub (Amersham, UK.) afin de lever les compressions.

20 L'analyse de séquence a été réalisée avec le logiciel DNASTAR (Madison, Wis., USA) et le programme de Genetic Computer Group (University of Wisconsin Biotechnology Center, Wis., USA) accessible en ligne sur INFOBIOGEN. Les recherches informatisées de similitude ont été réalisées avec le programme BLAST, les alignements multiples avec CLUSTAL V, et les arbres phylogénétiques ont été établis selon la méthode dite de "Neighbour-joining" (11).

30 4) Construction du recombinant exprimant l'ADN polymérase de *Thermococcus fumicolans*.

35 L'ADN polymérase de *Thermococcus fumicolans* ainsi que ses deux intéines, ont été exprimées en même temps chez *E. coli* avec le vecteur d'expression PARHS2 qui appartient à la famille des systèmes d'expression T7 (16) acquis auprès d'Eurogentec.

La PCR a été utilisée pour préparer le fragment complet de l'ADN polymérase et des deux intéines en

utilisant des amorces contenant les sites de restriction NdeI et BamHI :

- amorce Tfu Dir :

5'-TGG GGA TCC ATA TGA TCC TCG ATA CAG ACT ACA TC-3'

5

- amorce Tfu Rev :

5'-AAG CTT GGA TCC TCA TTT CTT CCC CAT TTT GAG CC-3'

10 Le mélange réactionnel contenait l'ADN polymérase GOLDSTAR (Eurogentec, B), l'enzyme Taq Extender (contenant la Pfu de Stratagene), le tampon d'extension avec les quatre dNTP (chacun à 0,2mM) et les amorces Tfu Dir et Tfu Rev à 50 pmoles dans un volume de 50µl final. L'amplification a été effectuée sur 20 cycles : 1 mn 94°C, 1 mn à 54°C et 6 mn à 72°C en utilisant un thermocycleur Stratagene 96-gradient. Les fragments de PCR ont ensuite été digérés par les enzymes NdeI et BamHI puis ligaturés aux mêmes sites du vecteur, rétablissant ainsi le codon d'initiation. La construction ainsi obtenue a été nommée PARHS2TFU1. Cette construction a été séquencée aux sites de jonction afin de vérifier son intégrité par rapport à la séquence de l'ADN génomique. Les tests d'expression ont été réalisés selon le protocole suivant: sélection des clones recombinants dans la souche *E. coli* Novablue, expression avec la souche BL21(DE3)pLySS dans un milieu 2xYT et induction de quatre heures à 1mM d'IPTG.

25 Les premiers essais d'induction ont été réalisés sur des cultures de 5 ml, induites ou non. Des précultures de nuit sont réalisées sans inducteur et relancées dans un milieu frais le matin (au dixième), jusqu'à ce que la densité optique (mesurée à 600nm) soit de 0,6, puis soit induites pendant 4 heures, soit non induites et arrêtées au bout de 2, 4 ou 14 heures. Quatre ml de cultures sont alors centrifugés à 4°C, 5000rpm pendant 10 minutes. Le culot est ensuite repris dans un tampon de lyse (Tris-HCl 10mM pH 7,5; NaCl 10mM, MgCl2 2 mM). Les cellules ainsi reprises sont alors lysées, soit avec du triton X-100 1% v/v, soit du lysozyme à 1mg/ml de

lyse et laissées sur glace environ 5 à 10 min. Les cellules sont ensuite thermodénaturées par une exposition de 20 min à 72°C. Ceci permet de détruire en grande partie les cellules de l'hôte mésophile sans détruire les protéines recombinantes. Le produit de lyse est ensuite centrifugé 20 min à 10 000rpm à 4°C. Le surnageant est récupéré afin de le tester en incorporation, en PCR ou sur gel. Les incorporations sont réalisées suivant deux techniques :

10 - Avec de la thymidine tritiée comme traceur.
L'incorporation est réalisée sur du thymus de veau activé (SIGMA Aldrich, F) dans le milieu de réaction suivant: Tris-HCl 50mM pH 8,8 ; DTT 1mM ; MgCl₂ 10mM ; KCl 10mM ; BSA 0,4 mg/ml; chaque dNTP à 0,4 mM.

15 - Avec du ³²P -dATP (Amersham) comme traceur.
L'incorporation est réalisée sur de l'ADN thymus de veau activé (Appligene) dans le milieu de réaction suivant: Tris-HCl pH 9 50mM; KCl 50mM; MgCl₂ 7mM; BSA 0,2mg/ml et (NH₄⁺)SO₄ (filtré) 16mM, avec un mélange des 4 dNTP à 500µM final chacun. Cette seconde méthode permet
20 d'estimer avec précision le nombre d'unités d'enzyme.

5) Purification.

25 a) Culture.

Après des essais en petits volumes, les cultures destinées à l'expression de l'enzyme recombinante ont été réalisées comme suit: production d'un inoculum de 700 ml (milieu 2x YT complémenté en
30 ampicilline et chloramphénicol) cultivé à 30°C jusqu'à DO = 0,8; ensemencement d'un fermenteur contenant 16 l du même milieu; culture pendant 4 h jusqu'à DO= 0,6, puis induction à l'IPTG 1 mM et culture pendant 4 h. La biomasse résultante est centrifugée 20 mn à 6 000 rpm à
35 4°C (Centrifugeuse JOUAN).

b) Lyse cellulaire et première étape de purification.

20 g de biomasse sont repris dans 80 ml de tampon (20 mM Tris-Cl pH 7,5; 10 mM NaCl; 2 mM MgCl₂; 1 mM EGTA; 1% Triton x100; 2,2 mM PMSF). Le mélange résultant, maintenu à 4°C maximum, est soniqué à 12 reprises successives (cycle de 15 s) jusqu'à obtention d'une solution liquide. Le surnageant est ensuite centrifugé 20 mn à 4°C à 20 000 rpm (SORVALL Ti45). Le surnageant est récupéré et traité par la chaleur (70°C pendant 10 mn) afin de thermodénaturer l'essentiel des protéines natives de *E. coli*, puis centrifugé à nouveau.

c) Chromatographie.

Les 70 ml de surnageant issus de l'étape précédente sont chargés sur une colonne Pharmacia Héparine Sépharose (30 ml de résine), après équilibration avec un tampon A (10 mM Tris-Cl pH 7,5; 0,5 mM EGTA; 5 mM MgCl₂; 10 mM β -mercaptoéthanol; 0,2% Triton x100 et 10% glycérol. Un lavage est effectué avec un tampon B (idem tampon A + 2 M NaCl) à raison de 0,3 ml/mn sur un système FPLC Pharmacia. Les différentes fractions sont récupérées en gradient de NaCl.

Les fractions actives ainsi récupérées sont dialysées pendant 5 h contre un tampon C: 10 mM Tris-Cl pH 7,5; 0,5 mM EGTA; 5 mM MgCl₂; 10 mM -mercaptoéthanol; 0,2% Triton x100; 10% glycérol; 50 mM NaCl. Les produits issus de dialyse sont successivement chargés sur une colonne d'affinité pour les protéines se liant à l'ADN (Pharmacia, Bleue HiTrap) et élués en gradient de NaCl avec les mêmes tampons que précédemment.

30ml de fractions actives obtenues précédemment sont chargées sur une colonne de phosphocellulose avec de la résine P11 de Whatmann (volume:20ml; diamètre: 2,5cm). Ces fractions ont été dialysées 5 heures contre le tampon suivant : KPO₄ pH7 20mM, EDTA 0,1mM, DTT 1mM, Glycérol 5%, Triton X-100 0,1% et KCl 0,1M. Le débit de

la colonne est réglé à 0,2ml/min, le tampon A de chargement est composé de KPO4 pH7 20mM, EDTA 0,1mM, DTT 1mM, Glycérol 5%, Triton X-100 0,1% et le gradient (entre 0% et 50% deB) est réalisé par le KCl présent dans tampon B à 2M.

Ayant déjà mis en évidence que cette polymérase ne se fixe pas sur une MonoQ ou une ressource Q et ce, quel que soit le pH utilisé, nous avons essayé de la récupérer en exclusion en faisant l'hypothèse d'une fixation de l'ADN contaminant par la colonne.

Tout d'abord un essai a été réalisé en sortie de la deuxième Bleue HiTrap avec un aliquot de 5ml et dialysé selon la même méthode que pour un passage sur une Bleue HiTrap.

Une deuxième tentative a été réalisée après la phosphocellulose et après deux dialyses des fractions les plus actives 45, 46 et 49. Les fractions sont tout d'abord chauffées 40 min à 85°C. Une première dialyse est alors réalisée contre le tampon KCl 0,1mM, K₂HPO₄ 1M pH7,5 pendant 3 heures. La deuxième dialyse est réalisée avec le tampon suivant : K₂HPO₄ pH 7,5 10mM, K₂PO₄ 10mM, KCl 25mM, DTT 1mM, Triton X-100 0,1%, Glycérol 10%, pendant 1 heure. La solution est alors chargée sur une colonne MonoQ à 0,5ml/min avec un gradient en NaCl de 0 à 20% (tampons utilisés pour l'Héparine).

6) Activités exonucléasiques.

Les tests d'exonuclease 3'-5' sont quantifiés par la libération de nucléotides marqués au ³²P.

A cette fin, une première étape permet de réaliser le marquage: l'ADN de est digéré par HindIII puis, le fragment de Klenow recopie l'ADN à partir des extrémités 3'-OH libres, dans un milieu contenant outre le tampon, l'ADN et l'enzyme, du ³²P dATP et du ³²P dTTP, les dGTPet dCTP étant froids. Après une heure à 37°C, les quatre dNTP froids sont rajoutés en excès pour une demi-

heure. La Klenow et les dNTP sont éliminés par une extraction au phénol et précipités à l'éthanol.

Les tests exonucléasiques sont effectués dans des solutions contenant les tampons des enzymes, 0,02mg/ml d'ADN marqué, et incubées toute la nuit à 72°C, 80°C et 95°C. Différents tampons contenant du MgCl₂ ou du MnSO₄ sont testés. Le même test est réalisé avec la Vent en témoin positif. 10l de solution de réaction sont alors déposés sur papier DE81 (Watmann), séchés puis comptés avant et après lavage (3 fois 5 min avec du Na₂HPO₄, 1 fois à l'eau puis à l'éthanol à 95%) en utilisant la technique de Cerenkov.

Pour le test d'activité exonucléasique 5'-3', la réaction de marquage utilise la polynucléotide kinase afin de marquer le substrat en 5'.

II - Résultats.

1) Isolement du gène de l'ADN polymérase de *Thermococcus fumicolans* ainsi que des deux gènes d'intéines.

L'ADN de *Thermococcus fumicolans*, digéré par une série d'enzymes de restriction, a été hybridé à des sondes de *P. furiosus* et *T. litoralis*, préparées par PCR et aux sondes de *Pyrococcus* sp. GE23 obtenues lors du clonage de cette autre ADN polymérase (Dépôt de brevet n°96 08631 auprès de l'INPI). Comme montré dans les figures 1a et 1b, l'hybridation de type Southern blot a révélé des fragments de deux types : un fragment HindIII-HindIII de 1,9 kb et un fragment XhoI-XhoI de 5 kb. Ces deux fragments ont été révélés uniquement avec la sonde ClaI-HindIII du gène de *Pyrococcus* sp. GE23, marquée au ³²P avec une exposition de deux heures. Les deux fragments repérés ont ensuite été récupérés et purifiés comme décrit précédemment dans Matériel et Méthodes puis clonés dans le vecteur pUC18 déphosphorylé et digéré par HindIII ou dans le vecteur pBluescript déphosphorylé et

digéré par XhoI. Environ 400 recombinants (*E. Coli* SURE) ont été criblés avec la sonde ClaI-HindIII. Sur ces 400 colonies, deux ont donné un signal positif lors de l'hybridation. (n° 9.26 et 12.79). Les deux clones ont été mis en culture en milieu LB-Amp et leurs profils de restriction étaient identiques, avec un insert de 1,8 kb. Le séquençage ultérieur de l'un de ces clones (désigné 557MACa) et la comparaison de séquence (Megalign, programme DNASTAR) ont montré qu'il correspond à la région promotrice et aux 1404 premières paires de base d'un gène d'une ADN polymérase appartenant à la famille B (2).

En ce qui concerne le fragment XhoI-XhoI de 5 kb, 700 recombinants ont été criblés sans aucun signal positif lors de l'hybridation. Cette absence de réussite pourrait être due à la présence d'une intéine au sein de ce fragment, le rendant alors instable dans un vecteur à haut nombre de copies de type pBluescript.

Après 12 heures d'exposition, un deuxième fragment HindIII-HindIII de 2 kb a été repéré par hybridation de type Southern sur la même membrane que précédemment, en utilisant comme sonde marquée la fin du gène de *Pyrococcus* sp. GE23 entre les sites XhoI et SalI (fragment obtenu par digestion de produit de PCR). Ce fragment a été cloné comme précédemment. Environ 200 clones recombinants ont été criblés et quatre d'entre eux ont donné un signal positif. Les quatre clones ont un profil identique après digestion par l'enzyme de restriction HindIII. L'un d'entre eux, 557MACc, a été séquencé et la comparaison de séquence a démontré qu'il s'agissait de la fin du gène de l'ADN polymérase précédemment identifiée.

Supposant que les fragments de gènes obtenus appartenaient à la même ADN polymérase, des oligonucléotides ont été utilisés afin d'amplifier la zone manquante. Ce fragment de PCR a ensuite été purifié sur gel comme décrit dans Matériel et Méthodes, et

utilisé comme sonde marquée radioactivement pour repérer sur la membrane le fragment manquant. Après deux heures de révélation un fragment HindIII-HindIII a été révélé, de 2 kb environ. Cloné comme précédemment, 600 colonies
5 ont alors été criblées, donnant quatre clones positifs. Les quatre clones avaient le même profil et l'un d'entre eux, nommé 557MACb a été séquencé et les comparaisons de séquences ont démontré qu'il s'agit de la partie intermédiaire de l'ADN polymérase et ce fragment,
10 délimité par deux sites HindIII, s'insère parfaitement entre 557MACa et 557MACc. L'ensemble de ces trois clones donne la séquence complète de l'ADN polymérase de *T. fumicolans*.

15 2) Position phylogénétique de *Thermococcus fumicolans*

Thermococcus fumicolans est une nouvelle espèce de Thermococcales décrite par Godfroy et al (8).

20 3) Séquences nucléotidiques et polypeptidiques de l'ADN polymérase de *Thermococcus fumicolans*

a) Séquences nucléotidiques

Les trois fragments délimités par des sites
25 HindIII ont été assemblés (figures 2 et 3). Ils forment à eux trois un fragment de 5039 paires de bases. Le premier fragment, issu du clone 557MACa, contient le codon de départ ATG en position 457 pb. La phase ouverte de lecture est alors ininterrompue sur 4572 paires de bases
30 jusqu'à rencontrer un codon STOP sur le fragment issu du clone 557MACc en position 5028 pb, six paires de bases avant le dernier site HindIII. Par alignement, avec les autres gènes d'ADN polymérases disponibles en banques (Pfu, Tli, GB-D, KOD, ...), avec la méthode CLUSTAL V,
35 complétée par un alignement manuel final nécessaire pour restituer les sites d'autoépissage des intéines, deux séquences d'insertions ont été mises en évidence.

- La première est insérée à la paire de base n° 1675 et se termine à 2754, étant ainsi répartie entre le fragment issu du clone 557MACa et celui du clone 557MACb.

5 - La deuxième est insérée à la base n° 3157 et se termine à 4323. Elle est, de même, répartie entre le clone 557MACb et 557MACc.

Ces deux séquences d'insertions forment, avec le reste de la séquence codant pour l'ADN polymérase, un seul cadre de lecture.

10

b) Séquences polypeptidiques.

La section codante du gène de l'ADN polymérase est donc constituée de trois parties disjointes. La première partie du gène, portée par 557MACa, comporte la zone codant pour l'exonucléase 3'-5', où après traduction, on reconnaît le motif FDIET. La deuxième partie, portée par 557MACb et la troisième partie portée par 557MACc comprennent les sites conservés SLYPSI, et YG.DTD.

20

Les deux séquences d'insertion sont situées sur des zones conservées de l'ADN polymérase, la première au site DFR/SLYPSII comme I-KOD-1 de *Pyrococcus* sp. KOD1 et la deuxième au site D/TDG comme I-Tli-1 de *T. litoralis*. Ces deux protéines sont libérées par autoépissage protéique. Elles comportent les sites de type LAGLIDADG répétés deux fois à 100 paires de bases de distance environ. Les alignements avec les autres intéines déposées dans les banques de séquences permettent de les assimiler à des endonucléases de restriction de type archaebactérien, endonucléases qui coupent l'ADN à l'endroit précis où leur gène s'insère.

25

30

c) Comparaison avec les autres séquences d'ADN polymérases.

35

L'alignement des différentes séquences polypeptidiques des ADN polymérases de Thermococcales disponibles en banque, *P. abyssi* strain GE5, *Pyrococcus*

sp. GE23, *Pyrococcus* sp. KOD1, *Pyrococcus* sp. GB-D, *P. furiosus*, *Pyrococcus* sp 9°N, *T. litoralis* avec la séquence de *T. fumicolans* (sans intéines), réalisé avec le programme CLUSTAL utilisant la matrice PAM250, donne les niveaux de similarité entre ces diverses polymérases figurant au tableau 1 ci-dessous.

Tableau 1

| Espèce | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------------|-----|------|------|------|------|------|------|------|
| 1 - <i>T. fumicolans</i> | xxx | 81,7 | 89,3 | 77,6 | 77,3 | 80,0 | 79,3 | 90,6 |
| 2 - <i>P. sp.</i> GB-D | | xxx | 81,3 | 85,1 | 76,9 | 89,0 | 88,1 | 82,6 |
| 3 - <i>P. sp.</i> KOD1 | | | xxx | 78,8 | 77,5 | 80,7 | 79,9 | 90,3 |
| 4 - <i>P. furiosus</i> | | | | xxx | 73,9 | 83,3 | 82,5 | 79,6 |
| 5 - <i>T. litoralis</i> | | | | | xxx | 75,9 | 75,0 | 76,5 |
| 6 - <i>P. sp.</i> GE23 | | | | | | xxx | 98,7 | 80,9 |
| 7 - <i>P. abyssi</i> | | | | | | | xxx | 80,2 |
| 8 - <i>P. sp.</i> 9°N | | | | | | | | xxx |

Le niveau de similarité le plus proche est celui observé avec *Pyrococcus* sp. 9°N (90,6%), ce qui indique clairement que l'ADN polymérase de *T. fumicolans* est originale au niveau de sa séquence et constitue de ce fait une nouvelle ADN polymérase.

d) Comparaison des intéines avec les autres intéines disponibles dans les banques.

Les séquences disponibles en banque ont été alignées selon les mêmes méthodes que précédemment. Les comparaisons de l'ensemble des intéines démontrent que celles-ci sont réparties en trois groupes correspondant aux sites d'insertion des motifs A (R/SLYPSI), B (KILAN/S) et C (D/TDG). L'analyse des niveaux de similarité et la recherche de relations phylogénétiques n'ont de sens que pour des intéines appartenant à la même classe, c'est à dire s'insérant dans un motif donné. Les niveaux de similarités pour I-Tfu-1 (classe A) et I-Tfu-2 (classe C) avec leurs intéines homologues déjà décrites sont donnés respectivement aux tableaux 2 et 3 ci-dessous.

Tableau 2

| | | 1 | 2 | 3 |
|---|---------|-----|------|------|
| 1 | I-Tfu-1 | xxx | 41,4 | 75,3 |
| 2 | I-Mja-1 | | xxx | 51,7 |
| 3 | I-KOD-1 | | | xxx |

Tableau 3

| | | 1 | 2 |
|---|------------|-----|------|
| 1 | I-Tfu-1 | xxx | 62,2 |
| 2 | I-I-Tfli-2 | | xxx |

5 I-Tfu-1 et I-Tfu-2 semblent représenter deux nouveaux "allèles" des "homing" endonucléases archaebactériennes des classes A et C.

I-Tfu-1 est le troisième allèle connu d'intéine s'insérant au site A des ADN polymérase d'Archaea, tandis que I-Tfu-2 est le second de sa classe.

10

5) Expression, caractérisation et activité de l'ADN polymérase de *T. fumicolans*.

a) Clonage et expression.

15

Un insert de 4595 pb obtenu par PCR longue-distance avec les amorces TfuDir et TfuRev et couvrant la totalité du gène de l'ADN polymérase de *Thermococcus fumicolans*, avec les deux intéines, a été cloné aux sites NdeI et BamHI d'un vecteur pour transformer la souche *E. coli* Novablue. Des mini-préparations d'ADN plasmidique ont été réalisées sur une dizaine de clones transformants et ont toutes donné un signal positif à l'hybridation. Deux clones ont été retenus sur la base de leur profil après une digestion par NdeI et SalI ou par HindIII. Ces deux plasmides ont ensuite été transformés dans la souche *E. coli* BL21(DE3)pLyss.

20

Des essais d'expression ont été réalisés en culture de 5 ml afin de déterminer les conditions optimales de culture et d'induction. Tout d'abord les essais de cultures ont été réalisés à 37°C, où l'on observe une lyse cellulaire trop importante, puis à 30°C où la culture ne lyse pratiquement pas. La culture est

30

réalisée en milieu 2xYT, où le rendement de production est meilleur et la lyse réduite, supplémenté avec de l'ampicilline (100µg/ml) et du chloramphenicol (15µg/ml). L'expression est alors induite en phase exponentielle de croissance (DO600nm = 0,6 à 0,7) avec une concentration en IPTG de 1mM, concentration qui s'est avérée optimale. Des échantillons sont prélevés avant induction puis 2 heures, 4 heures et 6 heures après induction, ainsi qu'après une nuit. Des prélèvements sont aussi réalisés sur les cultures non induites après 6, 8, 12 et 24 heures de culture.

Les échantillons sont alors traités comme indiqué au chapitre Matériel et Methodes, afin de tester le niveau d'activité de l'ADN polymérase recombinante par incorporation de thymidine tritiée ou de ³²P dCTP. Cette première technique, qui permet une approche qualitative, nous donne deux types de comportement. Un des clones est non inductible et la meilleure activité est obtenue après une nuit de culture (clone rTful-1). Un second clone est inductible et présente une activité maximale après quatre heures d'induction, activité qui diminue par la suite (clone rTful100-2). Deux autres clones, aussi testés sont faiblement inductibles et expriment très peu l'ADN polymérase.

Les clones rTful-1 et rTful100-2 sont ensuite testés en erlenmeyer de 50 ml dans les conditions décrites précédemment. Seul le clone inductible rTful100-2 a une expression constante en volume de 50 ml. La suite des travaux a donc été réalisée sur ce clone.

b) Fermentation et extraction des cellules.

La culture du clone Tful100-2 a été réalisée dans un fermenteur de 16 litres, dans le milieu 2xYT supplémenté en ampicilline et chloramphénicol. La préculture, de 750ml a été réalisée la veille et arrêtée en phase exponentielle (DO 600nm= 0,7-0,8) et laissée à 4°C pour la nuit. Le fermenteur a été préparé et mis en

température à 30°C avec le milieu de culture. La préculture a été remise à 30°C une heure avant son transfert dans le fermenteur. Le volume final de culture est de 15 litres. Les conditions sont les suivantes:

5 température = 30°C ; agitation = 300 rpm. L'induction avec 1 mM d'IPTG a été réalisée à DO 600 = 0,58. Le pH de la culture a été régulé à 7 pendant la phase d'acidification puis laissé libre lors de la phase

10 alcaline. Les bactéries ont été prélevées après quatre heures d'induction, alors que le pH était de 8,3. La culture a ensuite été centrifugée puis les cellules ont été divisées en trois lots. L'un d'eux, 20 g de pâte, a été repris dans 80 ml de tampon de lyse pour un

15 traitement ultérieur indiqué au chapitre Matériel et méthodes.

c) Purification de l'ADN polymérase de *T. fumicolans*.

La purification a été effectuée comme indiqué précédemment (Matériel et Méthodes). Pour la colonne Héparine -Sépharose, un gradient de 3 à 50% de tampon B, correspondant aux volume 365 ml à 1363 ml, permet de récupérer 73 fractions de 6 ml. Le pic d'activité de la polymérase est obtenu pour une valeur de gradient de 0,5

20 M environ, et correspond aux fractions 55/56 (dosées à 10 et 12 unités respectivement) comme indiqué à la figure 7. Ces fractions, incubées à 37°C pendant la nuit en présence d'ADN de pBR322, dégradent l'ADN et présentent en conséquence des traces de nucléase de l'hôte, non

25 visibles sur gel. Leur élimination, ou tout au moins une réduction substantielle de leur concentration a été réalisée par un passage sur colonne d'affinité (Bleue Hitrap).

30

Les fractions 54 à 60, regroupées et dialysées,

35 sont chargées à raison de 0,25 ml/mn. L'élution permet de récupérer 65 fractions de 5 ml avec des pics d'activité pour les fractions 36 à 56. Le dosage de

l'activité fait apparaître une concentration de 3 à 5 unités pour les fractions 36 à 40. Ces fractions, mises en présence d'ADN à 37°C, présentent une faible activité nucléasique.

5 Néanmoins, l'activité sur le plasmide pBR322 à 37°C toute la nuit montre une nette amélioration de la pureté de l'enzyme. Une deuxième colonne Bleue HiTrap a été réutilisée en prenant les fractions 36 à 44 et
10 dialysées comme précédemment. 25ml sur les 30ml de fraction ont été injectés sur la colonne avec un débit de 0,25ml/min (5 ml étant gardés pour essayer une MonoQ). Après cette deuxième colonne Bleue HiTrap, l'activité sur le plasmide pBR322 fermé et incubé toute la nuit avec des
15 fractions de la colonne, est nulle. L'incubation d'une heure à 72°C de fractions avec l'ADN de Lambda digéré par HindIII montre une très nette dégradation, mettant en évidence l'activité exonucléase 3'-5' associée à notre ADN polymérase. Le pic d'activité se situe entre les fractions 27 et 32. L'ADN polymérase plus pure est donc
20 sortie plus tôt sur le gradient de NaCl. Sur les fractions les plus actives, un comptage a été réalisé donnant une activité supérieure à 5U/ μ l pour les fractions 29, 30 et 31. La figure 4 montre le résultat sur gel SDS-PAGE. Suite à ces trois colonnes, la pureté
25 de la polymérase est nettement améliorée. Néanmoins, il reste des traces d'ADN de *E.coli* fixé à la polymérase et mises en évidence par PCR. Deux colonnes supplémentaires vont donc être mis en oeuvre. 30ml de fractions actives obtenues précédemment sont chargées sur une colonne de
30 phosphocellulose (celle-ci devrait fixer différemment l'ADN et la polymérase). L'activité de la polymérase est repérée par son activité exonucléasique 3'-5' élevée à 72°C sur l'ADN de Lambda digéré par HindIII. L'activité la plus forte se situe entre les fractions 40 et 49 avec
35 un pic net en 46 et 47. Sur ces fractions 40 à 49 l'ADN de pBR322 est intact après une nuit à 37°C. La figure 5 nous montre les résultats sur gel avec des activités

mesurées de 6U/ μ l pour la fraction 46, 4,5U/ μ l pour la 47 et 3U/ μ l pour la 48. Néanmoins il reste encore des traces d'ADN de *E. coli*.

5 Ayant déjà mis en évidence que cette polymérase ne se fixe pas sur une MonoQ ou une ressource Q et ce quel que soit le pH utilisé, nous avons essayé de la récupérer en exclusion en espérant une fixation de l'ADN contaminant par la colonne.

10 Tout d'abord un essai a été réalisé en sortie de la deuxième Bleue HiTrap avec 5ml dialysés. Les résultats ont été décevants, les fractions d'exclusion étant peu actives en incorporation.

15 Une deuxième tentative a été réalisée après la phosphocellulose et après deux dialyses des fractions les plus actives 45, 46 et 49. Les fractions sont tout d'abord chauffées 40 min à 85°C en raison d'une détection de contaminant dégradant le pBR322 à 72°C en une nuit. Aucune floculation n'est alors visible et l'extrait est mis sur glace. Un nouveau test démontre que la contamination semble réduite. Après les dialyses et le passage sur colonne, l'activité exonucléasique est mise en évidence sur les fractions d'exclusion 3 à 7. L'activité est dosée à 2U/ μ l et l'enzyme est visualisée sur la figure 10. Suite à cette dernière étape de purification, nous avons obtenu des résultats positifs en PCR et comparables à ceux obtenus pour la Vent.

d) Caractérisation des activités des fractions purifiées.

30 L'activité de l'ADN polymérase des différentes fractions a été dosée par incorporation au 32 P dATP selon le protocole décrit en Matériel et Méthodes.

- Amplification de gènes in vitro.

35 Un fragment de 459 pb a été amplifié à partir d'un ADN génomique d'Archaeobactérie (*Thermococcus* sp. GE

8) avec des amorces spécifiques. Différents tampons ont été utilisés:

- tampon 10x R: Tris HCl pH 8,8: 300mM; KCl: 500mM; MgCl₂: 30mM; Tween 20: 0,1%
- 5 - tampon 10x H: Tris HCl: pH 8,8: 300 mM; KCl: 500mM; MgCl₂: 15mM; Tween 20: 0,1%
- tampon 10x T: Tris HCl pH 8,8: 600mM; KCl: 500mM; MgCl₂: 15mM; Tween 20: 0,1%
- 10 - tampon 10x S: Tris Hcl pH 8,8: 200mM; KCl: 250mM; MgCl₂: 20mM; Tween 20: 0,1%

15 Trente cycles ont été réalisés, chacun comprenant une étape de dénaturation à 94°C pendant 30 sec., une étape d'hybridation des amorces à 51°C pendant 1 min. puis une étape d'élongation à 72°C pendant 2 min. La figure 8 présente les résultats obtenus avec un volume réactionnel de 50 µl pour des quantités d'ADN polymérase de *Thermococcus fumicolans* de 2,7 unités. En l'état actuel, les meilleurs résultats avec la Tfu sont obtenus avec le tampon R.

20 La figure 9 représente les résultats de l'amplification d'un fragment de 1,6 kb avec la Tfu purifiée sur colonnes d'héparine puis de séphacryl-bleue, et un tampon réactionnel 10x ayant la composition suivante: Tris HCl pH 8,8: 200mM; KCl: 100mM; (NH₄)₂SO₄: 25 100mM; MgSO₄: 20mM; Triton X-100: 1%.

- Activité exonucléasique.

30 Les tests d'activité, selon le protocole détaillé en Matériels et Méthodes, ne révèlent pas d'activité exonucléasique 5'-3' chez la Tfu. Ceci est en conformité avec la structure de l'enzyme déduite de l'analyse de la séquence polypeptidique qui ne fait pas apparaître de domaine fonctionnel exonucléase 5'-3', contrairement à ce qui est observé pour DNA polI de *E. coli* et la Taq.

35 Les tests d'activité, selon le protocole détaillé au chapitre Matériels et Méthodes, font

apparaître une activité exonucléasique 3'-5' (activité proof-reading ou de correction d'erreurs) chez la Tfu, à un niveau sensiblement égal à celui de la Vent comme montré dans le tableau 4 ci-dessous. Le tableau 4 rapporte la mesure de l'activité exonucléasique 3'-5' de la Tfu en fonction de la concentration en dNTP, et en comparaison avec la Vent.

Tableau 4

| Conc. dNTP (mM) | Tfu | Témoin + (Vent) | Témoin - (Pab exo-) |
|--------------------|--------|--------------------|------------------------|
| 0 | 30 000 | 25 000 | 100 000 |
| 0,1 | 40 000 | - | |
| 0,2 | 40 000 | 56 000 | |
| 0,3 | 40 000 | 58 000 | |
| 0,4 | 40 000 | 70 000 | |
| 0,5 | - | - | |
| 0,75 | 40 000 | 84 000 | |
| 1 | 90 000 | 106 000 | 100 000 |

Ces résultats sont en conformité avec la structure de l'enzyme déduite de l'analyse de la séquence polypeptidique qui révèle la présence d'un domaine exonucléasique 3'-5' en position N-terminale, ainsi que la présence des motifs catalytiques caractéristiques de ce domaine. La Tfu est sensible à une concentration de l'ordre de 0,8 mM de dNTP, tandis que la Vent manifeste une sensibilité dès 0,5 mM. Cette activité est connue pour améliorer *in vitro* la fidélité des polymérases utilisées en PCR.

En outre, cette activité exonucléasique est confirmée par un test plus simple. Les fractions purifiées, dépourvues d'activité nucléasique à 37°C, mises en présence d'ADN de digéré par HindIII puis exposé à 72°C pendant la nuit, dégradent complètement cet ADN, mettant ainsi en évidence la présence de l'activité 3'-5' exonucléase de la Tfu ainsi que sa thermostabilité.

- Thermostabilité.

La thermostabilité mesurée selon le protocole décrit précédemment (Matériel et Méthodes), ou mieux, l'activité résiduelle en incorporation à 72°C après exposition de l'enzyme à des températures élevées pendant des temps variables, est donnée dans le tableau 5 ci-dessous.

Tableau 5

| Température (°C) | Demi-vie polymérase | Activité exo après 3h |
|------------------|---------------------|-----------------------|
| 92 | - | 100% |
| 95 | 4 h | |
| 100 | 2 h | |

Cette thermostabilité, inférieure à celle des polymérases issues d'organismes plus hyperthermophiles tels que les *Pyrococcus*, n'en demeure pas moins très largement supérieure à celles des polymérases issues des *Thermus* et en particulier toutes les Taq. La thermostabilité de l'enzyme recombinante purifiée, tant pour le domaine polymérase que pour l'exonucléase, est de toute manière très largement supérieure à tous les besoins connus en PCR.

REFERENCES BIBLIOGRAPHIQUES.

1. Birnboim, H. C., and J. Doly. 1979. A rapid extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*. 7:1503.
2. Braithwaite, D. K., and J. Ito. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research*. 21(4):787-802.
3. Charbonnier, F. 1993. Paris Sud.
4. Chong, S. R., Y. Shao, H. Paulus, J. Benner, F. B. Perler, and M. Q. Xu. 1996. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein - The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *Journal of Biological Chemistry*. 271(36):22159-22168.
5. Davis, E. O., S. G. Sedgwick, and J. Colston. 1991. Novel structure of *Mycobacterium tuberculosis* implies processing of the gene product. *Journal of Bacteriology*. 173(18):5653-5662.
6. Davis, E. O., S. G. Sedgwick, and M. J. Colston. 1991. Novel structure of the *recA* locus of *Mycobacterium tuberculosis* implies processing of the gene product. *Journal of Bacteriology*. 173(18):5653-5662.
7. Fsihi, H., V. Vincent, and S. T. Cole. 1996. Homing events in the *gyrA* gene of some mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America*. 93(8):3410-3415.
8. Godfroy, A., J. R. Meunier, J. Guézennec, F. Lesongeur, G. Raguénès, A. Rimbault, and G. Barbier. 1996. *Thermococcus fumicolans* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent in the North Fiji Basin. *International Journal of Systematic Bacteriology*. 46(4):1113-1119.
9. Hodges, R. A., F. B. Perler, C. J. Noren, and W. E. Jack. 1992. Protein Splicing Removes

Intervening Sequences in an Archaea DNA Polymerase.
Nucleic Acids Res. 20(23):6153-6157.

10. Kong, H. M., R. B. Kucera, and W. E. Jack. 1993. Characterization of a DNA Polymerase from the Hyperthermophile Archaea *Thermococcus litoralis* - Vent DNA Polymerase, Steady State Kinetics, Thermal Stability, Processivity, Strand Displacement, and Exonuclease Activities. J Biol Chem. 268(3):1965-1975.
11. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4(4):406-425.
12. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
13. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5467-5473.
14. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of molecular biology. 98:503.
15. Southworth, M. W., H. Kong, R. B. Kucera, J. Ware, H. W. Jannasch, and F. B. Perler. 1996. Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus* sp. 9°N-7 and mutations affecting 3'-5' exonuclease activity. Proc. Ntl. Acad. Sci. USA. 93:5281-5285.
16. Studier, F. W., A. H. Rosenberg, F. J. Dunn, and J. W. Dubendorff. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185:60-89.
17. Sutherland, K. J., C. M. Henneke, P. Towner, and D. W. Hough. 1990. Citrate synthase from the thermophilic archaeobacterium *Thermoplasma acidophilum*. Cloning and sequencing of the gene. European Journal of Biochemistry. 194:839-844.

18. Uemori, T., Y. Ishino, H. Toh, K. Asada, and I. Kato. 1993. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Research. 21(2):259-265.

LISTE DE SÉQUENCES

(1) INFORMATION GÉNÉRALES :

- (i) DEPOSANT : APPLIGENE - ONCOR
- (ii) TITRE DE L'INVENTION: ADN POLYMERASE
THERMOSTABLE D'ARCHAEBACTÉRIES DE L'ESPECE
THERMOCOCCUS fumicolans
- (iii) NOMBRE DE SEQUENCES: 4

(2) INFORMATION POUR LA SEQ ID NO:1 :

- (i) CARACTERISTIQUES DE LA SEQUENCE:
 - (A) LONGUEUR: 5039 paires de bases
 - (B) TYPE: nucléotide
 - (C) NOMBRE DE BRIN: double
 - (D) CONFIGURATION: linéaire
- (ii) TYPE DE MOLECULE: ADN
- (ix) CARACTERISTIQUES
 - (A) NOM/CLE: séquence codante de l'ADN
polymérase de *THERMOCOCCUS fumicolans* Tfu
 - (B) EMPLACEMENT: de 457 à 5028
- (ix) CARACTERISTIQUES
 - (A) NOM/CLE: séquence codante de l'intéine
I-Tfu-1
 - (B) EMPLACEMENT: de 1675 à 2754
- (ix) CARACTERISTIQUES
 - (A) NOM/CLE: séquence codante de l'intéine
I-Tfu-2
 - (B) EMPLACEMENT: de 3157 à 4323
- (ix) CARACTERISTIQUES
 - (A) NOM/CLE: codon stop
 - (B) EMPLACEMENT: de 5026 à 5028
- (xi) DESCRIPTION DE LA SEQUENCES: SEQ ID NO:1 :

| | |
|---|-------------------------|
| AGCTTAAAGC GTCCGCCACT ACTTCCTGAA AGCTCACGCG GTAAAACAGC TCCATGCTCG | 60 |
| GCTCTTCGAT GGGAGGTTTA AAAAGGTGGT GGTGAGGTTT ATTAGGAAGA AGGCTCAACT | 120 |
| AGAGACGGTG GGAGTATGGA AGAGGTCGAC AGGCTCGTGT TCAACTTTCC CCTCTTCAAA | 180 |
| GATTACTGGG AAAAGGAGCG GTTCCTCAAG GTCGTTGGGC TTCTGGTGAG CCACCAGATA | 240 |
| ACGTTTGAGA AAGCTGCCGA GCTTCTGGAC ATGAGGCTCG AAGAGCTGGC GTTCCTCCTT | 300 |
| GACAAGCTCG GCGTTGAGTA CTCGCTTCTT GATGATGAAG AGGCCAGACT TGAGAGAGAA | 360 |
| GAGGCCAATA AGCTCATGGG GGAAATGAAG GGTGGAGCGT TTGTCTGATT CTTCTGAGCT | 420 |
| GTTATTGGTG TTTCACAGGC TGGGAGGTGG TGGATT ATG ATC CTC GAT ACA GAC | 474 |
| | Met Ile Leu Asp Thr Asp |
| | 1 5 |
| TAC ATC ACC GAA GAC GGA AGG CCC GTC ATC AGG GTG TTC AAG AAG GAG | 522 |
| Tyr Ile Thr Glu Asp Gly Arg Pro Val Ile Arg Val Phe Lys Lys Glu | |
| 10 15 20 | |

| | |
|---|------|
| AAC GGC GAG TTC AAA ATC GAG TAC GAC AGG GAC TTC GAG CCT TAC ATC Asn Gly Glu Phe Lys Ile Glu Tyr Asp Arg Asp Phe Glu Pro Tyr Ile 25 30 35 | 570 |
| TAC GCT CTC CTG AAG GAC GAT TCC GCG ATC GAG GAC GTC AAG AAG ATA Tyr Ala Leu Leu Lys Asp Asp Ser Ala Ile Glu Asp Val Lys Lys Ile 40 45 50 | 618 |
| ACT GCA AGC CGG CAC GGT ACC ACC GTC AGG GTC GTC AGG GCC GGG AAG Thr Ala Ser Arg His Gly Thr Thr Val Arg Val Val Arg Ala Gly Lys 55 60 65 70 | 666 |
| GTG AAG AAG AAG TTC CTC GGC AGG CCG ATA GAG GTC TGG AAG CTC TAC Val Lys Lys Lys Phe Leu Gly Arg Pro Ile Glu Val Trp Lys Leu Tyr 75 80 85 | 714 |
| TTC ACC CAT CCC CAG GAC GTT CCG GCA ATC AGG GAC AAA ATC AGG GAG Phe Thr His Pro Gln Asp Val Pro Ala Ile Arg Asp Lys Ile Arg Glu 90 95 100 | 762 |
| CAC CCT GCC GTG GTC GAC ATA TAT GAG TAC GAC ATA CCC TTT GCC AAG His Pro Ala Val Val Asp Ile Tyr Glu Tyr Asp Ile Pro Phe Ala Lys 105 110 115 | 810 |
| CGC TAC CTC ATC GAT AAG GGC CTC ATC CCG ATG GAG GGC GAC GAG GAG Arg Tyr Leu Ile Asp Lys Gly Leu Ile Pro Met Glu Gly Asp Glu Glu 120 125 130 | 858 |
| CTC AAG ATG CTC GCC TTC GAC ATC GAG ACG CTC TAC CAC GAG GGC GAG Leu Lys Met Leu Ala Phe Asp Ile Glu Thr Leu Tyr His Glu Gly Glu 135 140 145 150 | 906 |
| GAG TTC GCC GAG GGG CCT ATT CTT ATG ATA AGC TAT GCC GAC GAG GAA Glu Phe Ala Glu Gly Pro Ile Leu Met Ile Ser Tyr Ala Asp Glu Glu 155 160 165 | 954 |
| GGG GCG AGG GTA ATA ACC TGG AAG AAG ATC GAC CTT CCC TAC GTT GAC Gly Ala Arg Val Ile Thr Trp Lys Lys Ile Asp Leu Pro Tyr Val Asp 170 175 180 | 1002 |
| GTC GTT TCA ACG GAG AAG GAG ATG ATA AAG CGC TTC CTG AAG GTT GTC Val Val Ser Thr Glu Lys Glu Met Ile Lys Arg Phe Leu Lys Val Val 185 190 195 | 1050 |
| AAG GAG AAG GAC CCC GAT GTC CTC ATA ACC TAC AAC GGC GAC AAC TTC Lys Glu Lys Asp Pro Asp Val Leu Ile Thr Tyr Asn Gly Asp Asn Phe 200 205 210 | 1098 |
| GAC TTC GCT TAC CTC AAG AAG CGC TCC GAG AAG CTC GGC GTT AAG TTC Asp Phe Ala Tyr Leu Lys Lys Arg Ser Glu Lys Leu Gly Val Lys Phe 215 220 225 230 | 1146 |
| ATC CTC GGA AGG GAC GGC AGC GAG CCG AAG ATA CAG AGG ATG GGC GAC Ile Leu Gly Arg Asp Gly Ser Glu Pro Lys Ile Gln Arg Met Gly Asp 235 240 245 | 1194 |
| CGC TTC GCC GTC GAG GTG AAG GGA AGA ATA CAC TTC GAC CTC TAC CCC Arg Phe Ala Val Glu Val Lys Gly Arg Ile His Phe Asp Leu Tyr Pro 250 255 260 | 1242 |

| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| GTC | ATA | AGA | CAC | ACC | ATC | AAC | CTG | CCC | ACC | TAC | ACG | CTG | GAG | GCC | GTC | 1290 |
| Val | Ile | Arg | His | Thr | Ile | Asn | Leu | Pro | Thr | Tyr | Thr | Leu | Glu | Ala | Val | |
| | | 265 | | | | | 270 | | | | | 275 | | | | |
| TAC | GAG | GCG | ATT | TTT | GGG | CAG | CCA | AAG | GAG | AAG | GTC | TAC | GCT | GAG | GAG | 1338 |
| Tyr | Glu | Ala | Ile | Phe | Gly | Gln | Pro | Lys | Glu | Lys | Val | Tyr | Ala | Glu | Glu | |
| | 280 | | | | | 285 | | | | | 290 | | | | | |
| ATA | GCG | CAG | GCC | TGG | GAA | ACG | GGC | GAG | GGG | CTT | GAG | CGC | GTC | GCG | CGC | 1386 |
| Ile | Ala | Gln | Ala | Trp | Glu | Thr | Gly | Glu | Gly | Leu | Glu | Arg | Val | Ala | Arg | |
| 295 | | | | | 300 | | | | | 305 | | | | | 310 | |
| TAC | TCG | ATG | GAG | GAC | GCC | AAG | GTA | ACC | TAC | GAG | CTG | GGA | AGG | GAG | TTC | 1434 |
| Tyr | Ser | Met | Glu | Asp | Ala | Lys | Val | Thr | Tyr | Glu | Leu | Gly | Arg | Glu | Phe | |
| | | | | 315 | | | | | 320 | | | | | 325 | | |
| TTC | CCG | ATG | GAG | GCC | CAA | CTT | TCT | CGG | CTG | GTC | GGT | CAG | AGC | TTC | TGG | 1482 |
| Phe | Pro | Met | Glu | Ala | Gln | Leu | Ser | Arg | Leu | Val | Gly | Gln | Ser | Phe | Trp | |
| | | | 330 | | | | | 335 | | | | | 340 | | | |
| GAC | GTC | TCG | CGC | TCC | AGC | ACC | GGC | AAC | CTC | GTC | GAG | TGG | TAC | CTC | CTC | 1530 |
| Asp | Val | Ser | Arg | Ser | Ser | Thr | Gly | Asn | Leu | Val | Glu | Trp | Tyr | Leu | Leu | |
| | | 345 | | | | | 350 | | | | | 355 | | | | |
| AGG | AAG | GCC | TAC | GAG | AGG | AAC | GAG | CTG | GCA | CCG | AAC | AAG | CCC | TCC | GGC | 1578 |
| Arg | Lys | Ala | Tyr | Glu | Arg | Asn | Glu | Leu | Ala | Pro | Asn | Lys | Pro | Ser | Gly | |
| | 360 | | | | | 365 | | | | | 370 | | | | | |
| AGA | GAA | CTT | GAG | AGG | CGC | CGC | GGG | GGC | TAC | GCC | GGC | GGC | TAC | GTC | AAG | 1626 |
| Arg | Glu | Leu | Glu | Arg | Arg | Arg | Gly | Gly | Tyr | Ala | Gly | Gly | Tyr | Val | Lys | |
| 375 | | | | | 400 | | | | | 405 | | | | | 410 | |
| GAG | CCG | GAG | AGG | GGA | CTT | TGG | GAG | AAC | ATA | GCT | TAT | TTA | GAT | TTT | AGG | 1674 |
| Glu | Pro | Glu | Arg | Gly | Leu | Trp | Glu | Asn | Ile | Ala | Tyr | Leu | Asp | Phe | Arg | |
| | | | | 415 | | | | | 420 | | | | | 425 | | |
| TGT | CAT | CCT | GCC | GAC | ACT | AAA | GTC | ATT | GTC | AAA | GGG | AAG | GGC | GTT | GTA | 1722 |
| Cys | His | Pro | Ala | Asp | Thr | Lys | Val | Ile | Val | Lys | Gly | Lys | Gly | Val | Val | |
| | | | 430 | | | | 435 | | | | | | 440 | | | |
| AAC | ATC | AGC | GAA | GTT | AGG | GAG | GGG | GAC | TAC | GTT | CTC | GGC | ATA | GAC | GGC | 1770 |
| Asn | Ile | Ser | Glu | Val | Arg | Glu | Gly | Asp | Tyr | Val | Leu | Gly | Ile | Asp | Gly | |
| | | 445 | | | | | 450 | | | | | 455 | | | | |
| TGG | CAG | AAG | GTT | CAA | AGG | GTC | TGG | GAG | TAT | GAT | TAC | GAG | GGA | GAA | CTC | 1818 |
| Trp | Gln | Lys | Val | Gln | Arg | Val | Trp | Glu | Tyr | Asp | Tyr | Glu | Gly | Glu | Leu | |
| | 460 | | | | | 465 | | | | | 470 | | | | | |
| GTA | AAT | ATA | AAC | GGC | CTT | AAG | TGC | ACA | CCG | AAC | CAT | AAG | CTT | CCG | GTC | 1866 |
| Val | Asn | Ile | Asn | Gly | Leu | Lys | Cys | Thr | Pro | Asn | His | Lys | Leu | Pro | Val | |
| 475 | | | | | 480 | | | | | 485 | | | | | 490 | |
| GTT | AGG | AGG | ACT | GAG | AGG | CAG | ACT | GCG | ATA | AGG | GAC | AGC | CTT | GCA | AAG | 1914 |
| Val | Arg | Arg | Thr | Glu | Arg | Gln | Thr | Ala | Ile | Arg | Asp | Ser | Leu | Ala | Lys | |
| | | | | 495 | | | | 500 | | | | | | 505 | | |
| TCT | TTT | CTC | ACG | AAA | AAA | GTT | AAA | GGT | AAG | CTG | ATA | ACC | ACG | CCT | CTC | 1962 |
| Ser | Phe | Leu | Thr | Lys | Lys | Val | Lys | Gly | Lys | Leu | Ile | Thr | Thr | Pro | Leu | |
| | | | 510 | | | | | 515 | | | | | 520 | | | |

| | |
|---|------|
| TTT GAA AAA ATC GGG AAG ATC GAG CGA GAG GAC GTG CCA GAA GAG GAG | 2010 |
| Phe Glu Lys Ile Gly Lys Ile Glu Arg Glu Asp Val Pro Glu Glu Glu | |
| 525 530 535 | |
| ATA CTC AAA GGA GAA CTC GCC GGA ATA ATC CTG GCT GAG GGC ACA CTC | 2058 |
| Ile Leu Lys Gly Glu Leu Ala Gly Ile Ile Leu Ala Glu Gly Thr Leu | |
| 540 545 550 | |
| CTG AGA AAG GAT GTC GAG TAC TTT GAC TCT TCC AGA GGG AAG AAG AGA | 2106 |
| Leu Arg Lys Asp Val Glu Tyr Phe Asp Ser Ser Arg Gly Lys Lys Arg | |
| 555 560 565 570 | |
| GTA TCA CAC CAG TAC AGG GTT GAA ATA ACC GTT GGG GCG CAG GAG GAG | 2154 |
| Val Ser His Gln Tyr Arg Val Glu Ile Thr Val Gly Ala Gln Glu Glu | |
| 575 580 585 | |
| GAC TTC CAG AGG AGG ATC GTT TAC ATT TTC GAA CGC CTC TTT GGG GTA | 2202 |
| Asp Phe X Arg Arg Ile Val Tyr Ile Phe Glu Arg Leu Phe Gly Val | |
| 590 595 600 | |
| ACT CCC ACT GTT TAC CGG AAA AAG AAC ACA AAC GCA ATA ACG TTC AAA | 2250 |
| Thr Pro Ser Val Tyr Arg Lys Lys Asn Thr Asn Ala Ile Thr Phe Lys | |
| 605 610 615 | |
| GTT GCC AAA AAA GAG GTT TAT CTT AGG GTT AGG GAA ATT ATG GAT GGC | 2298 |
| Val Ala Lys Lys Glu Val Tyr Leu Arg Val Arg Glu Ile Met Asp Gly | |
| 620 625 630 | |
| ATT GAG AAC CTC CAC GCT CCT TCT GTG TTA AGG GGC TTT TTT GAA GGA | 2346 |
| Ile Glu Asn Leu His Ala Pro Ser Val Leu Arg Gly Phe Phe Glu Gly | |
| 635 640 645 650 | |
| GAC GGA AGC GTC AAC AAG GTC CGG AAG ACA GTG GTA GTG AAT CAG GGC | 2394 |
| Asp Gly Ser Val Asn Lys Val Arg Lys Thr Val Val Val Asn Gln Gly | |
| 655 660 665 | |
| ACC AAT AAT GAA TGG AAA ATT GAA GTG GTG TCA AAA CTC CTC AAC AAG | 2442 |
| Thr Asn Asn Glu Trp Lys Ile Glu Val Val Ser Lys Leu Leu Asn Lys | |
| 670 675 680 | |
| TTG GGG ATT CCG CAT AGA AGG TAC ACA TAC GAT TAC ACC GAA AGA GAA | 2490 |
| Leu Gly Ile Pro His Arg Arg Tyr Thr Tyr Asp Tyr Thr Glu Arg Glu | |
| 685 690 695 | |
| AAA ACC ATG ACA ACG CAT ATA CTT GAG ATA GCC GGC AGG GAT GGC TTA | 2538 |
| Lys Thr Met Thr His Ile Leu Glu Ile Ala Gly Arg Asp Gly Leu | |
| 700 705 710 | |
| ATC CTT TTC CAG ACC ATT GTG GGA TTC ATA AGC ACT GAG AAG AAC ATG | 2586 |
| Ile Leu Phe Gln Thr Ile Val Gly Phe Ile Ser Thr Glu Lys Asn Met | |
| 715 720 725 730 | |
| GCG CTG GAG GAG GCA ATC AGG AAC AGG GAA GTG AAC CGC CTA GAA AAC | 2634 |
| Ala Leu Glu Glu Ala Ile Arg Asn Arg Glu Val Asn Arg Leu Glu Asn | |
| 735 740 745 | |
| AAT GCC TTC TAT ACC CTA GCC GAC TTT ACG GCG AAG ACA GAG TAC TAC | 2682 |
| Asn Ala Phe Tyr Thr Leu Ala Asp Phe Thr Ala Lys Thr Glu Tyr Tyr | |
| 750 755 780 | |

| | | | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|-------------------|------------|------|
| AAG Lys | GGC Gly | AAA Lys | GTT Val | TAC Tyr | GAC Asp | TTA Leu | ACC Thr | CTT Leu | GAG Glu | GGA Gly | ACG Thr | CCC Pro | TAT Tyr | TAC Tyr | TTC Phe | 2730 |
| | | 785 | | | | | 790 | | | | | 795 | | | | |
| GCC Ala | AAT Asn | GGC Gly | ATA Ile | CTG Leu | ACC Thr | CAC His | AAT Asn | <u>TCG</u> Ser | CTA Leu | TAT Tyr | CCT Pro | TCG Ser | ATT Ile | ATA Ile | ATT Ile | 2778 |
| | | 800 | | | | | 805 | | | | 810 | | | | | |
| TCC Ser | CAC His | AAC Asn | GTC Val | TCC Ser | CCC Pro | GAT Asp | ACG Thr | CTC Leu | AAC Asn | CGC Arg | GAG Glu | GGC Gly | TGC Cys | GGG Gly | GAG Glu | 2826 |
| | | 815 | | | 820 | | | | | 825 | | | | | 830 | |
| TAC Tyr | GAC Asp | GAG Glu | GCT Ala | CCG Pro | CAG Gln | GTA Val | GGG Gly | CAT His | CGC Arg | TTT Phe | TGT Cys | AAG Lys | GAC Asp | TTC Phe | CCC Pro | 2874 |
| | | | | 835 | | | | | 840 | | | | | 845 | | |
| GGC Gly | TTC Phe | ATC Ile | CCC Pro | AGC Ser | CTC Leu | CTC Leu | GGT Gly | GAC Asp | CTG Leu | CTC Leu | GAC Asp | GAG Glu | AGG Arg | CAG Gln | AAG Lys | 2922 |
| | | | 855 | | | | | 860 | | | | | 865 | | | |
| GTA Val | AAG Lys | AAG Lys | CAC His | ATG Met | AAG Lys | GCC Ala | ACG Thr | GTG Val | GAC Asp | CCG Pro | ATA Ile | GAG Glu | AAG Lys | AAG Lys | CTC Leu | 2970 |
| | | | 870 | | | | 875 | | | | | 880 | | | | |
| CTC Leu | GAT Asp | TAC Tyr | AGG Arg | CAG Gln | CGC Arg | GCA Ala | ATT Ile | AAA Lys | ATC Ile | CTC Leu | GCC Ala | AAC Asn | AGC Ser | TTC Phe | TAC Tyr | 3018 |
| | | 885 | | | | 890 | | | | | 895 | | | | | |
| GGC Gly | TAC Tyr | TAT Tyr | GGC Gly | TAC Tyr | GCA Ala | AAG Lys | GCC Ala | CGC Arg | TGG Trp | TAC Tyr | TGC Cys | AAG Lys | GAG Glu | TGC Cys | GCC Ala | 3066 |
| | | 900 | | | 905 | | | | | 910 | | | | | 915 | |
| GAG Glu | AGC Ser | GTT Val | ACC Thr | GCC Ala | TGG Trp | GGC Gly | AGG Arg | CAG Gln | TAC Tyr | ATT Ile | GAG Glu | ACC Thr | ACC Thr | ATG Met | AGG Arg | 3114 |
| | | | | 920 | | | | | 925 | | | | | 930 | | |
| GAA Glu | ATA Ile | GAG Glu | GAA Glu | AAA Lys | TTT Phe | GGC Gly | TTT Phe | AAA Lys | GTG Val | CTG Leu | TAC Tyr | GCG Ala | GAT Asp | <u>AGT</u> Ser | GTT Val | 3162 |
| | | | 935 | | | | | 940 | | | | | 945 | | | |
| ACA Thr | GGG Gly | GAC Asp | ACA Thr | GAG Glu | GTA Val | ACC Thr | ATC Ile | AGA Arg | AGA Arg | AAC Asn | GGC Gly | AGG Arg | ATT Ile | GAG Glu | TTC Phe | 3210 |
| | | 950 | | | | | 955 | | | | | 960 | | | | |
| GTT Val | CCA Pro | ATC Ile | GAG Glu | AAA Lys | CTC Leu | TTT Phe | GAG Glu | CGC Arg | GTT Val | GAT Asp | CAC His | CGT Arg | GTT Val | GGT Gly | GAG Glu | 3258 |
| | | 965 | | | | 970 | | | | | 975 | | | | | |
| AAG Lys | GAG Glu | TAC Tyr | TGC Cys | GTT Val | CTT Leu | GGA Gly | GGG Gly | GTT Val | GAG Glu | GCA Ala | CTG Leu | ACA Thr | CTC Leu | GAC Asp | AAC Asn | 3306 |
| | | 980 | | | 985 | | | | 990 | | | | | 995 | | |
| AGG Arg | GGC Gly | AGG Arg | CTC Leu | GTG Val | TGG Trp | AAG Lys | AAG Lys | GTT Val | CCG Pro | TAC Tyr | GTC Val | ATG Met | AGA Arg | CAT His | AAA Lys | 3354 |
| | | | | 1000 | | | | | 1005 | | | | | 1010 | | |
| ACG Thr | GAC Asp | AAA Lys | AGA Arg | ATC Ile | TAT Tyr | AGG Arg | GTA Val | TGG Trp | TTC Phe | ACC Thr | AAC Asn | TCT Ser | TGG Trp | TAC Tyr | CTT Leu | 3402 |
| | | | 1015 | | | | | 1020 | | | | | | 1025 | | |

| | |
|---|------|
| GAC GTG ACG GAG GAT CAC TCG CTA ATA GGC TAC CTG AAC ACA AGC AAA Asp Val Thr Glu Asp His Ser Leu Ile Gly Tyr Leu Asn Thr Ser Lys 1030 1035 1040 | 3450 |
| GTC AAA CCC GGA AAG CCC TTG AAA GAG CGT CTC GTC GAG GTC AAG CCA Val Lys Pro Gly Lys Pro Leu Lys Glu Arg Leu Val Glu Val Lys Pro 1045 1050 1055 | 3498 |
| GAA GAA TTG GGG GGT AAG GTC AAG TCT CTC ATT ACG CCC AAT CGG CCA Glu Glu Leu Gly Gly Lys Val Lys Ser Leu Ile Thr Pro Asn Arg Pro 1060 1065 1070 1075 | 3546 |
| ATT GCC CGT ACC ATC AAG GCC AAC CCC ATT GCC GTC AAG CTC TGG GAG Ile Ala Arg Thr Ile Lys Ala Asn Pro Ile Ala Val Lys Leu Trp Glu 1080 1085 1090 | 3594 |
| TTA ATT GGC CTG CTG GTG GGA GAT GGC AAC TGG GGT GGA CAA TCG AAC Leu Ile Gly Leu Leu Val Gly Asp Gly Asn Trp Gly Gly Gln Ser Asn 1095 1100 1105 | 3642 |
| TGG GCC AAA TAC TAC GTT GGC CTC TCC TGT GGG CTG GAT AAA GCC GAA Trp Ala Lys Tyr Tyr Val Gly Leu Ser Cys Gly Leu Asp Lys Ala Glu 1110 1115 1120 | 3690 |
| ATA GAG AGA AAA GTC CTG AAC CCT TTA AGA GAG GCA AGC GTC ATC TCC Ile Glu Arg Lys Val Leu Asn Pro Leu Arg Glu Ala Ser Val Ile Ser 1125 1130 1135 | 3738 |
| AAC TAC TAC GAC AAG AGC AAG AAG GGC GAC GTT TCC ATA CTC TCC AAG Asn Tyr Tyr Asp Lys Ser Lys Lys Gly Asp Val Ser Ile Leu Ser Lys 1140 1145 1150 1155 | 3786 |
| TGG CTC GCC GGA TTC ATG GTC AAA TAC TTC AAA GAT GAA AAT GGG AAC Trp Leu Ala Gly Phe Met Val Lys Tyr Phe Lys Asp Glu Asn Gly Asn 1160 1165 1170 | 3834 |
| AAG GCC ATT CCC AGC TTC ATG TTC AAC CTT CCA AGG GAA TAC ATA GAG Lys Ala Ile Pro Ser Phe Met Phe Asn Leu Pro Arg Glu Tyr Ile Glu 1175 1180 1185 | 3882 |
| GCC TTT CTA CGG GGG CTG TTT TCA GCG GAC GGA ACG GTA AGC TTG CGT Ala Phe Leu Arg Gly Leu Phe Ser Ala Asp Gly Thr Val Ser Leu Arg 1190 1195 1200 | 3930 |
| AGA GGA ATC CCA GAA ATT AGA CTG ACA AGC GTT AAC AGA GAG CTT AGT Arg Gly Ile Pro Glu Ile Arg Leu Thr Ser Val Asn Arg Glu Leu Ser 1205 1210 1215 | 3978 |
| GAT GCC GTG AGA AAG TTG CTG TGG CTG GTT GGG GTC TCC AAC TCA CTA Asp Ala Val Arg Lys Leu Leu Trp Leu Val Gly Val Ser Asn Ser Leu 1220 1225 1230 1235 | 4026 |
| TTC ACC GAA ACC AAG CCA AAC CGG TAC CTG GAG AAA GAA AGT GGA ACG Phe Thr Glu Thr Lys Pro Asn Arg Tyr Leu Glu Lys Glu Ser Gly Thr 1240 1245 1250 | 4074 |
| CAT TCG ATT CAC GTG AGG ATA AAG AAC AAG CAT CGC TTT GCC GAT AGA His Ser Ile His Val Arg Ile Lys Asn Lys His Arg Phe Ala Asp Arg 1255 1260 1265 | 4122 |

| | |
|---|------|
| ATA GGC TTT CTC ATA GAC AGA AAA TCC ACC AAA CTC TCC GAG AAC CTG Ile Gly Phe Leu Ile Asp Arg Lys Ser Thr Lys Leu Ser Glu Asn Leu 1270 1275 1280 | 4170 |
| GGG GGA CAT ACA AAC AAG AAG AGG GCT TAC AAA TAT GAT TTT GAC TTG Gly Gly His Thr Asn Lys Lys Arg Ala Tyr Lys Tyr Asp Phe Asp Leu 1285 1290 1295 | 4218 |
| GTA TAC CCC AGA AAA ATC GAA GAG ATA ACC TAC GAC GGC TAC GTC TAT Val Tyr Pro Arg Lys Ile Glu Glu Ile Thr Tyr Asp Gly Tyr Val Tyr 1300 1305 1310 1315 | 4266 |
| GAC ATC CAG GTT GAG GGA ACC CAC AGG TTC TTC GCC AAC GGA ATA CTC Asp Ile Glu Val Glu Gly Thr His Arg Phe Phe Ala Asn Gly Ile Leu 1320 1325 1330 | 4314 |
| GTT CAC AAC ACA GAC GGC TTT TTC GCA ACA ATC CCC GGA GCG GAC GCC Val His Asn Thr Asp Gly Phe Phe Ala Thr Ile Pro Gly Ala Asp Ala 1335 1340 1345 | 4362 |
| GAG ACG GTC AAA AAG AAG GCC AGG GAG TTC CTT AAC TAC ATT AAC CCC Glu Thr Val Lys Lys Lys Ala Arg Glu Phe Leu Asn Tyr Ile Asn Pro 1350 1355 1360 | 4410 |
| AAG CTG CCC GGT CTC CTC GAA CTC GAG TAC GAG GGC TTC TAC AGG CGC Lys Leu Pro Gly Leu Leu Glu Leu Glu Tyr Glu Gly Phe Tyr Arg Arg 1365 1370 1375 | 4458 |
| GGT TTC TTC GTA ACC AAG AAG AAG TAC GCG GTG ATA GAC GAG GAG GGC Gly Phe Phe Val Thr Lys Lys Lys Tyr Ala Val Ile Asp Glu Glu Gly 1380 1385 1390 1395 | 4506 |
| AAG ATA ACG ACG CGC GGG CTT GAG ATC GTC CGG CGC GAC TGG AGT GAG Lys Ile Thr Thr Arg Gly Leu Glu Ile Val Arg Arg Asp Trp Ser Glu 1400 1405 1410 | 4554 |
| GTG GCT AAG GAG ACG CAG GCG AGG GTC TTG GAG GCC ATA CTG CGG CAC Val Ala Lys Glu Thr Gln Ala Arg Val Leu Glu Ala Ile Leu Arg His 1415 1420 1425 | 4602 |
| GGT GAC GTC GAG GAG GCC GTG AGG ATT GTC AAG GAA GTG ACG GAA AAG Gly Asp Val Glu Glu Ala Val Arg Ile Val Lys Glu Val Thr Glu Lys 1430 1435 1440 | 4650 |
| CTG AGC AAG TAC GAG GTT CCG CCA GAG AAG CTC GTC ATC CAC GAG CAG Leu Ser Lys Tyr Glu Val Pro Pro Glu Lys Leu Val Ile His Glu Gln 1445 1450 1455 | 4698 |
| ATT ACC AGG GAG CTG AAG GAC TAC AAG GCC ACC GGC CCG CAC GTG GCC Ile Thr Arg Glu Leu Lys Asp Tyr Lys Ala Thr Gly Pro His Val Ala 1460 1465 1470 1475 | 4746 |
| ATA GCG AAG CGC CTC GCC GCG AGG GGG ATT AAG GTT CGC CCT GGG ACA Ile Ala Lys Arg Leu Ala Ala Arg Gly Ile Lys Val Arg Pro Gly Thr 1480 1485 1490 | 4794 |
| GTC ATC AGC TAC ATC GTC CTG AAA GGT TCC GGC AGG ATA GGG GAC AGG Val Ile Ser Tyr Ile Val Leu Lys Gly Ser Gly Arg Ile Gly Asp Arg 1495 1500 1505 | 4842 |

ACG ATA CCC TTC GAC GAG TTC GAC CCC ACG AAG CAC AGG TAC GAT GCG 4890
 Thr Ile Pro Phe Asp Glu Phe Asp Pro Thr Lys His Arg Tyr Asp Ala
 1510 1515 1520

GAG TAC TAC ATC GAG AAC CAG GTT CTG CCG GCG GTG GAG AGA ATC CTC 4938
 Glu Tyr Tyr Ile Glu Asn Gln Val Leu Pro Ala Val Glu Arg Ile Leu
 1525 1530 1535

AAG GCC TTC GGC TAC AAG AAG GAG GAT TTG CGC TAC CAG AAG ACG CGG 4986
 Lys Ala Phe Gly Tyr Lys Lys Glu Asp Leu Arg Tyr Gln Lys Thr Arg
 1540 1545 1550 1555

CAG GTT GGG CTG GGG GCG TGG CTC AAA ATG GGG AAG AAA TGA 5028
 Gln Val Gly Leu Gly Ala Trp Leu Lys Met Gly Lys Lys
 1560 1565 1568

AGGCCAAGCT T 5039

(2) INFORMATION POUR LA SEQ ID NO:2 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 774 acides aminés

(ii) TYPE DE MOLECULE: protéine

(ix) CARACTERISTIQUES

(A) NOM/CLE: ADN polymérase de *THERMOCOCCUS fumicolans* Tfu

(xi) DESCRIPTION DE LA SEQUENCES: SEQ ID NO:2 :

Met Ile Leu Asp Thr Asp Tyr Ile Thr Glu Asp Gly Arg Pro Val Ile
 1 5 10 15

Arg Val Phe Lys Lys Glu Asn Gly Glu Phe Lys Ile Glu Tyr Asp Arg
 20 25 30

Asp Phe Glu Pro Tyr Ile Tyr Ala Leu Leu Lys Asp Asp Ser Ala Ile
 35 40 45

Glu Asp Val Lys Lys Ile Thr Ala Ser Arg His Gly Thr Thr Val Arg
 50 55 60

Val Val Arg Ala Gly Lys Val Lys Lys Lys Phe Leu Gly Arg Pro Ile
 65 70 75 80

Glu Val Trp Lys Leu Tyr Phe Thr His Pro Gln Asp Val Pro Ala Ile
 85 90 95

Arg Asp Lys Ile Arg Glu His Pro Ala Val Val Asp Ile Tyr Glu Tyr
 100 105 110

Asp Ile Pro Phe Ala Lys Arg Tyr Leu Ile Asp Lys Gly Leu Ile Pro
 115 120 125

Met Glu Gly Asp Glu Glu Leu Lys Met Leu Ala Phe Asp Ile Glu Thr
 130 135 140

Leu Tyr His Glu Gly Glu Glu Phe Ala Glu Gly Pro Ile Leu Met Ile
 145 150 155 160

Ser Tyr Ala Asp Glu Glu Gly Ala Arg Val Ile Thr Trp Lys Lys Ile
 165 170 175

Asp Leu Pro Tyr Val Asp Val Val Ser Thr Glu Lys Glu Met Ile Lys
 180 185 190
 Arg Phe Leu Lys Val Val Lys Glu Lys Asp Pro Asp Val Leu Ile Thr
 195 200 205
 Tyr Asn Gly Asp Asn Phe Asp Phe Ala Tyr Leu Lys Lys Arg Ser Glu
 210 215 220
 Lys Leu Gly Val Lys Phe Ile Leu Gly Arg Asp Gly Ser Glu Pro Lys
 225 230 235 240
 Ile Gln Arg Met Gly Asp Arg Phe Ala Val Glu Val Lys Gly Arg Ile
 245 250 255
 His Phe Asp Leu Tyr Pro Val Ile Arg His Thr Ile Asn Leu Pro Thr
 260 265 270
 Tyr Thr Leu Glu Ala Val Tyr Glu Ala Ile Phe Gly Gln Pro Lys Glu
 275 280 285
 Lys Val Tyr Ala Glu Glu Ile Ala Gln Ala Trp Glu Thr Gly Glu Gly
 290 295 300
 Leu Glu Arg Val Ala Arg Tyr Ser Met Glu Asp Ala Lys Val Thr Tyr
 305 310 315 320
 Glu Leu Gly Arg Glu Phe Phe Pro Met Glu Ala Gln Leu Ser Arg Leu
 325 330 335
 Val Gly Gln Ser Phe Trp Asp Val Ser Arg Ser Ser Thr Gly Asn Leu
 340 345 350
 Val Glu Trp Tyr Leu Leu Arg Lys Ala Tyr Glu Arg Asn Glu Leu Ala
 355 360 365
 Pro Asn Lys Pro Ser Gly Arg Glu Leu Glu Arg Arg Arg Gly Gly Tyr
 370 375 380
 Ala Gly Gly Tyr Val Lys Glu Pro Glu Arg Gly Leu Trp Glu Asn Ile
 385 390 395 400
 Ala Tyr Leu Asp Phe Arg Ser Leu Tyr Pro Ser Ile Ile Ile Ser His
 405 410
 Asn Val Ser Pro Asp Thr Leu Asn Arg Glu Gly Cys Gly Glu Tyr Asp
 415 420 425
 Glu Ala Pro Gln Val Gly His Arg Phe Cys Lys Asp Phe Pro Gly Phe
 430 435 440
 Ile Pro Ser Leu Leu Gly Asp Leu Leu Asp Glu Arg Gln Lys Val Lys
 445 450 455
 Lys His Met Lys Ala Thr Val Asp Pro Ile Glu Lys Lys Leu Leu Asp
 460 465 470 475
 Tyr Arg Gln Arg Ala Ile Lys Ile Leu Ala Asn Ser Phe Tyr Gly Tyr
 480 485 490

Tyr Gly Tyr Ala Lys Ala Arg Trp Tyr Cys Lys Glu Cys Ala Glu Ser
 495 500 505
 Val Thr Ala Trp Gly Arg Gln Tyr Ile Glu Thr Thr Met Arg Glu Ile
 510 515 520
 Glu Glu Lys Phe Gly Phe Lys Val Leu Tyr Ala Asp Thr Asp Gly Phe
 525 530 535
 Phe Ala Thr Ile Pro Gly Ala Asp Ala Glu Thr Val Lys Lys Lys Ala
 540 545 555 560
 Arg Glu Phe Leu Asn Tyr Ile Asn Pro Lys Leu Pro Gly Leu Leu Glu
 565 570 575
 Leu Glu Tyr Glu Gly Phe Tyr Arg Arg Gly Phe Phe Val Thr Lys Lys
 580 585 590
 Lys Tyr Ala Val Ile Asp Glu Glu Gly Lys Ile Thr Thr Arg Gly Leu
 595 600 605
 Glu Ile Val Arg Arg Asp Trp Ser Glu Val Ala Lys Glu Thr Gln Ala
 610 615 620
 Arg Val Leu Glu Ala Ile Leu Arg His Gly Asp Val Glu Glu Ala Val
 625 630 635 640
 Arg Ile Val Lys Glu Val Thr Glu Lys Leu Ser Lys Tyr Glu Val Pro
 645 650 655
 Pro Glu Lys Leu Val Ile His Glu Gln Ile Thr Arg Glu Leu Lys Asp
 660 665 670
 Tyr Lys Ala Thr Gly Pro His Val Ala Ile Ala Lys Arg Leu Ala Ala
 675 680 685
 Arg Gly Ile Lys Val Arg Pro Gly Thr Val Ile Ser Tyr Ile Val Leu
 690 695 700
 Lys Gly Ser Gly Arg Ile Gly Asp Arg Thr Ile Pro Phe Asp Glu Phe
 705 710 715 720
 Asp Pro Thr Lys His Arg Tyr Asp Ala Glu Tyr Tyr Ile Glu Asn Gln
 725 730 735
 Val Leu Pro Ala Val Glu Arg Ile Leu Lys Ala Phe Gly Tyr Lys Lys
 740 745 750
 Glu Asp Leu Arg Tyr Gln Lys Thr Arg Gln Val Gly Leu Gly Ala Trp
 755 760 765
 Leu Lys Met Gly Lys Lys
 770 774

(2) INFORMATION POUR LA SEQ ID NO:3 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 360 acides aminés

(ii) TYPE DE MOLECULE: protéine

(ix) CARACTERISTIQUES

(A) NOM/CLE: intéine I-Tfu-1

(xi) DESCRIPTION DE LA SEQUENCES: SEQ ID NO:3 :

Cys His Pro Ala Asp Thr Lys Val Ile Val Lys Gly Lys Gly Val Val
 1 5 10 15
 Asn Ile Ser Glu Val Arg Glu Gly Asp Tyr Val Leu Gly Ile Asp Gly
 20 25 30
 Trp Gln Lys Val Gln Arg Val Trp Glu Tyr Asp Tyr Glu Gly Glu Leu
 35 40 45
 Val Asn Ile Asn Gly Leu Lys Cys Thr Pro Asn His Lys Leu Pro Val
 50 55 60
 Val Arg Arg Thr Glu Arg Gln Thr Ala Ile Arg Asp Ser Leu Ala Lys
 65 70 75 80
 Ser Phe Leu Thr Lys Lys Val Lys Gly Lys Leu Ile Thr Thr Pro Leu
 85 90 95
 Phe Glu Lys Ile Gly Lys Ile Glu Arg Glu Asp Val Pro Glu Glu Glu
 100 105 110
 Ile Leu Lys Gly Glu Leu Ala Gly Ile Ile Leu Ala Glu Gly Thr Leu
 115 120 125
 Leu Arg Lys Asp Val Glu Tyr Phe Asp Ser Ser Arg Gly Lys Lys Arg
 130 135 140
 Val Ser His Gln Tyr Arg Val Glu Ile Thr Val Gly Ala Gln Glu Glu
 145 150 155 160
 Asp Phe X Arg Arg Ile Val Tyr Ile Phe Glu Arg Leu Phe Gly Val
 165 170 175
 Thr Pro Ser Val Tyr Arg Lys Lys Asn Thr Asn Ala Ile Thr Phe Lys
 180 185 190
 Val Ala Lys Lys Glu Val Tyr Leu Arg Val Arg Glu Ile Met Asp Gly
 195 200 205
 Ile Glu Asn Leu His Ala Pro Ser Val Leu Arg Gly Phe Phe Glu Gly
 210 215 220
 Asp Gly Ser Val Asn Lys Val Arg Lys Thr Val Val Val Asn Gln Gly
 225 230 235 240
 Thr Asn Asn Glu Trp Lys Ile Glu Val Val Ser Lys Leu Leu Asn Lys
 245 250 255
 Leu Gly Ile Pro His Arg Arg Tyr Thr Tyr Asp Tyr Thr Glu Arg Glu
 260 265 270

Lys Thr Met Thr Thr His Ile Leu Glu Ile Ala Gly Arg Asp Gly Leu
 275 280 285
 Ile Leu Phe Gln Thr Ile Val Gly Phe Ile Ser Thr Glu Lys Asn Met
 290 295 300
 Ala Leu Glu Glu Ala Ile Arg Asn Arg Glu Val Asn Arg Leu Glu Asn
 305 310 315 320
 Asn Ala Phe Tyr Thr Leu Ala Asp Phe Thr Ala Lys Thr Glu Tyr Tyr
 325 330 335
 Lys Gly Lys Val Tyr Asp Leu Thr Leu Glu Gly Thr Pro Tyr Tyr Phe
 340 345 350
 Ala Asn Gly Ile Leu Thr His Asn
 355 360

(2) INFORMATION POUR LA SEQ ID NO:4 :

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 389 acides aminés

(ii) TYPE DE MOLECULE: protéine

(ix) CARACTERISTIQUES

(A) NOM/CLE: intéine I-Tfu-2

(xi) DESCRIPTION DE LA SEQUENCES: SEQ ID NO:4 :

Ser Val Thr Gly Asp Thr Glu Val Thr Ile Arg Arg Asn Gly Arg Ile
 1 5 10 15
 Glu Phe Val Pro Ile Glu Lys Leu Phe Glu Arg Val Asp His Arg Val
 20 25 30
 Gly Glu Lys Glu Tyr Cys Val Leu Gly Gly Val Glu Ala Leu Thr Leu
 35 40 45
 Asp Asn Arg Gly Arg Leu Val Trp Lys Lys Val Pro Tyr Val Met Arg
 50 55 60
 His Lys Thr Asp Lys Arg Ile Tyr Arg Val Trp Phe Thr Asn Ser Trp
 65 70 75 80
 Tyr Leu Asp Val Thr Glu Asp His Ser Leu Ile Gly Tyr Leu Asn Thr
 85 90 95
 Ser Lys Val Lys Pro Gly Lys Pro Leu Lys Glu Arg Leu Val Glu Val
 100 105 110
 Lys Pro Glu Glu Leu Gly Gly Lys Val Lys Ser Leu Ile Thr Pro Asn
 115 120 125
 Arg Pro Ile Ala Arg Thr Ile Lys Ala Asn Pro Ile Ala Val Lys Leu
 130 135 140
 Trp Glu Leu Ile Gly Leu Leu Val Gly Asp Gly Asn Trp Gly Gly Gln
 145 150 155 160
 Ser Asn Trp Ala Lys Tyr Tyr Val Gly Leu Ser Cys Gly Leu Asp Lys
 165 170 175

Ala Glu Ile Glu Arg Lys Val Leu Asn Pro Leu Arg Glu Ala Ser Val
 180 185 190
 Ile Ser Asn Tyr Tyr Asp Lys Ser Lys Lys Gly Asp Val Ser Ile Leu
 195 200 205
 Ser Lys Trp Leu Ala Gly Phe Met Val Lys Tyr Phe Lys Asp Glu Asn
 210 215 220
 Gly Asn Lys Ala Ile Pro Ser Phe Met Phe Asn Leu Pro Arg Glu Tyr
 225 230 235 240
 Ile Glu Ala Phe Leu Arg Gly Leu Phe Ser Ala Asp Gly Thr Val Ser
 245 250 255
 Leu Arg Arg Gly Ile Pro Glu Ile Arg Leu Thr Ser Val Asn Arg Glu
 260 265 270
 Leu Ser Asp Ala Val Arg Lys Leu Leu Trp Leu Val Gly Val Ser Asn
 275 280 285
 Ser Leu Phe Thr Glu Thr Lys Pro Asn Arg Tyr Leu Glu Lys Glu Ser
 290 295 300
 Gly Thr His Ser Ile His Val Arg Ile Lys Asn Lys His Arg Phe Ala
 305 310 315 320
 Asp Arg Ile Gly Phe Leu Ile Asp Arg Lys Ser Thr Lys Leu Ser Glu
 325 330 335
 Asn Leu Gly Gly His Thr Asn Lys Lys Arg Ala Tyr Lys Tyr Asp Phe
 340 345 350
 Asp Leu Val Tyr Pro Arg Lys Ile Glu Glu Ile Thr Tyr Asp Gly Tyr
 355 360 365
 Val Tyr Asp Ile Glu Val Glu Gly Thr His Arg Phe Phe Ala Asn Gly
 370 375 380
 Ile Leu Val His Asn
 385 389

REVENDICATIONS

- 5 1) ADN polymérase purifiée thermostable
d'archabactéries de l'espèce *Thermococcus fumicolans*
ayant un poids moléculaire de l'ordre de 89 000 daltons
et ses dérivés enzymatiquement équivalents.
- 10 2) ADN polymérase selon la revendication 1 dont
la séquence en acides aminés est représentée dans la
liste de séquence en annexe sous le numéro SEQ ID NO:1 ou
un fragment de celle-ci ou encore un assemblage de tels
fragments.
- 15 3) ADN polymérase selon la revendication 2 dont
la séquence en acides aminés est représentée dans la
liste de séquence en annexe sous le numéro SEQ ID NO:2.
- 20 4) Une séquence d'ADN constituée par ou
comprenant la séquence codant pour une ADN polymérase
selon quelconque des revendications 1 à 3.
- 25 5) Une séquence d'ADN selon la revendication 4
constituée par ou comprenant la séquence comprise entre
les nucléotides 357 à 5028 de la séquence SED ID NO: 1,
ou un fragment de celle-ci ou encore un assemblage de
tels fragments.
- 30 6) Une séquence d'ADN selon l'une des
revendications 4 à 5 constituée par ou comprenant les
nucléotides 357 à 1674 et 2755 à 3156 et 4324 à 5028 de
la séquence d'ADN représentée dans la liste de séquences
en annexe sous le numéro SED ID NO: 1.
- 35 7) Un vecteur contenant la séquence d'ADN de
l'une quelconque des revendications 4 à 6.

8) Un hôte transformé par un vecteur selon la revendication 7.

5 9) Procédé de préparation d'une ADN polymérase thermostable d'archaebactéries de l'espèce *Thermococcus fumicolans*, caractérisé en ce que l'on cultive l'hôte selon la revendication 8 dans des conditions permettant l'expression de ladite ADN polymérase et en ce que l'on extrait et récupère celle-ci par tout moyen approprié.

10

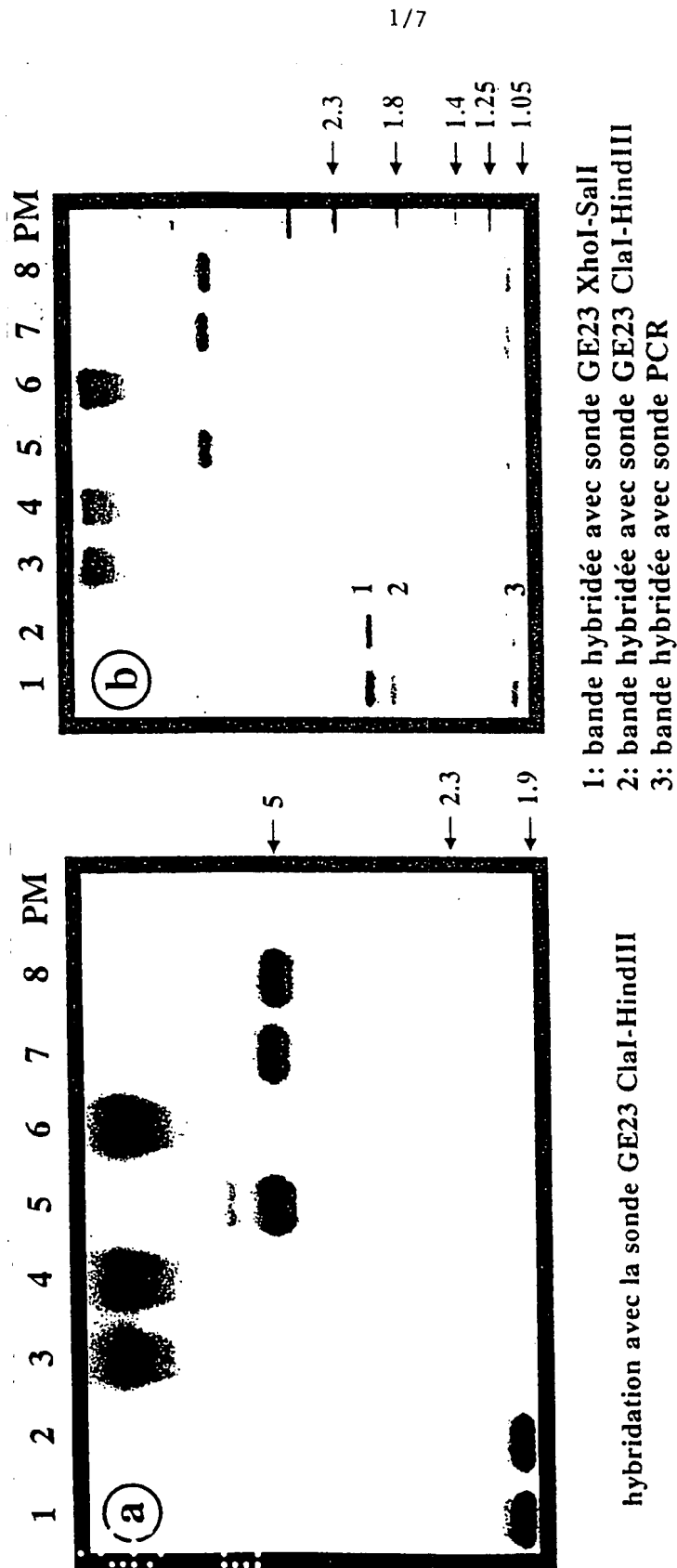
10) Procédé d'amplification enzymatique d'une séquence d'acide nucléique caractérisé en ce que l'on met en oeuvre une ADN polymérase thermostable selon l'une quelconque des revendications 1 à 3.

15

11) Intéine purifiée thermostable d'archaebacteries de l'espèce *Thermococcus fumicolans*.

20

Fig.1



hybridation avec la sonde GE23 ClaI-HindIII

Enzymes de restriction: utilisées:

1: HindIII-HindIII; 2: HindIII-XbaI; 3: PstI-PstI; 4: PstI-XbaI; 5: PstI-XhoI; 6: XbaI-XbaI; 7: XbaI-XhoI; 8: XhoI-XhoI;
PM= marqueur de poids moléculaire

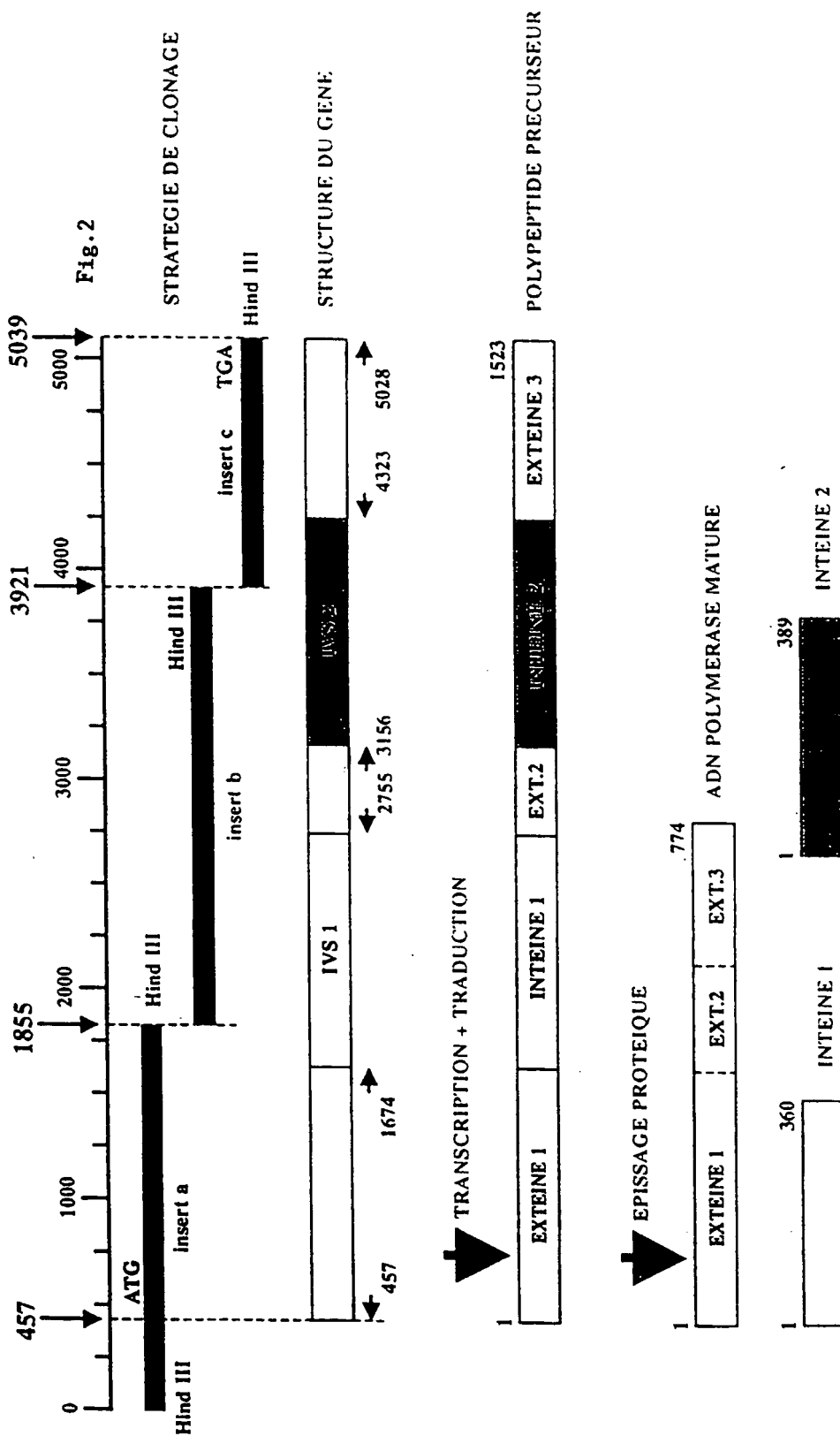
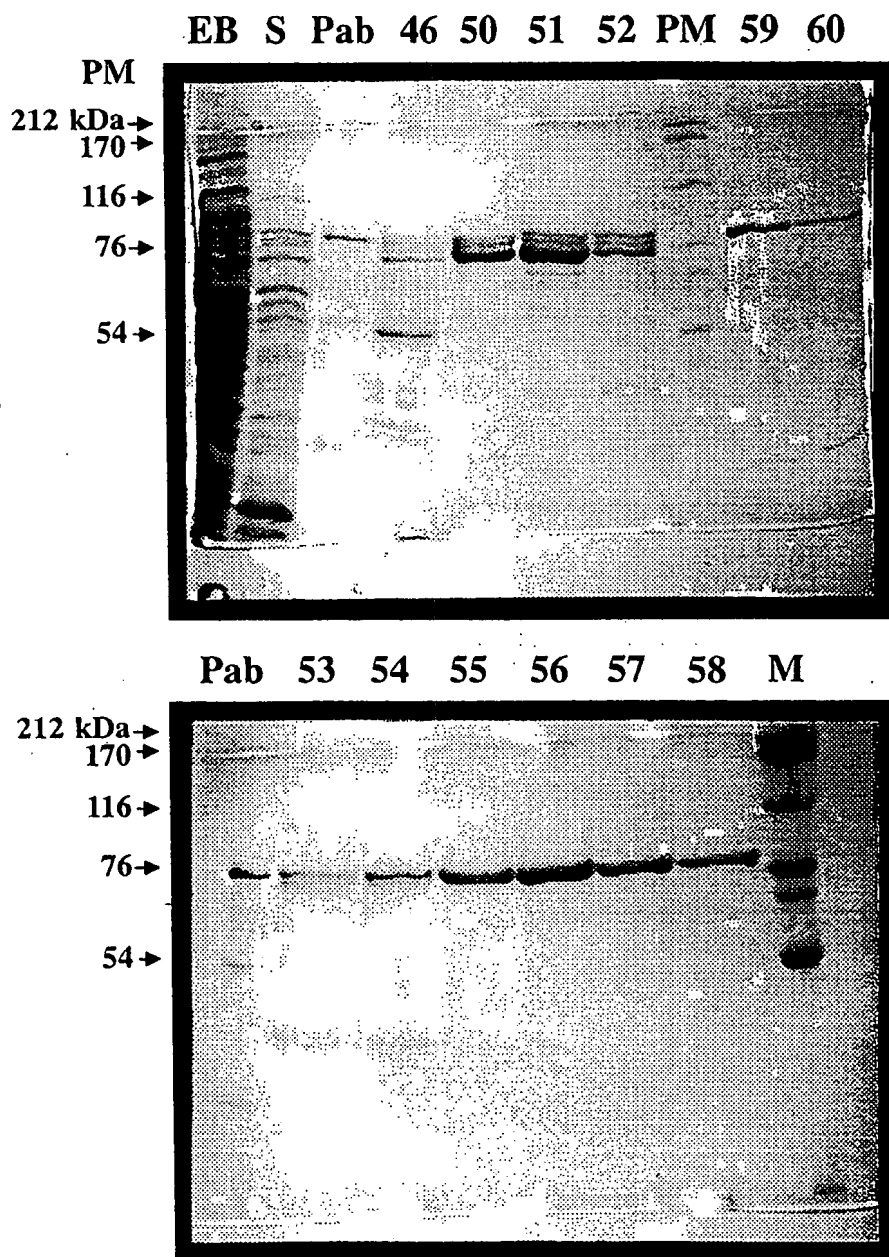


Fig.3



EB = extrait brut de sonication

S = surnageant chauffé

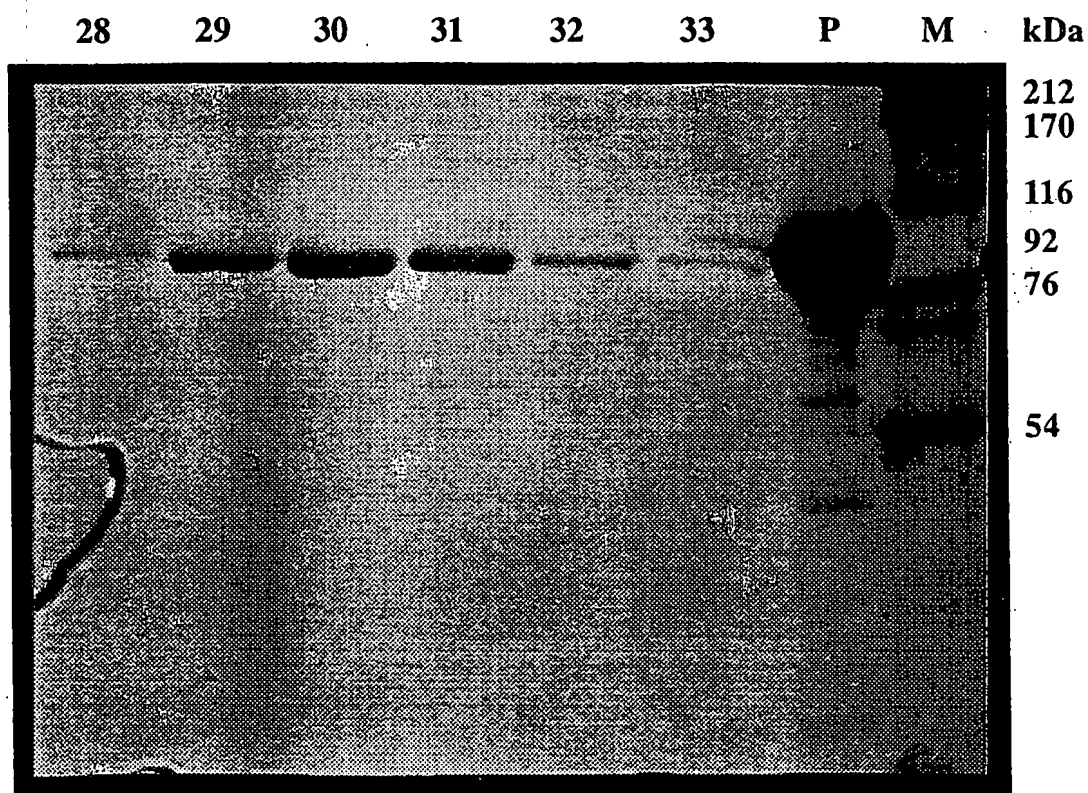
Pab \approx 50U de Pab

PM = poids moléculaire en kDa

FEUILLE DE REMPLACEMENT (REGLE 26)

4/7

Fig.4

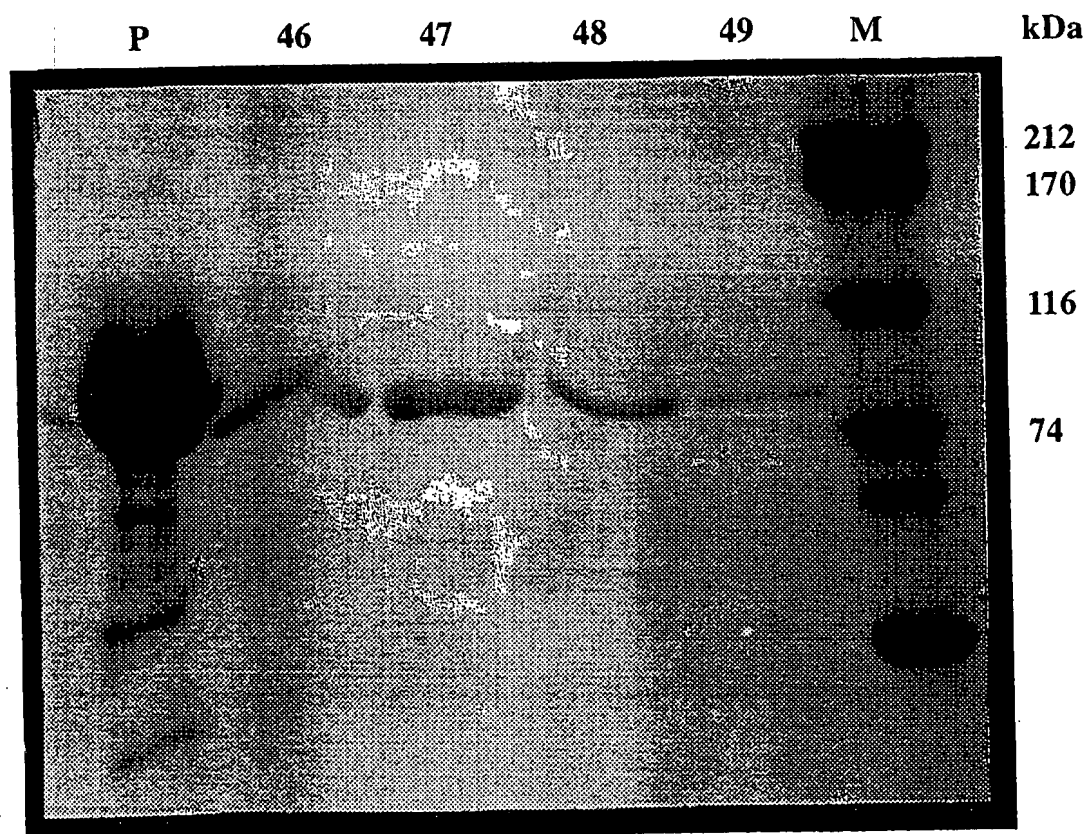


28, 29, 30, 31, 32, 33: fractions après la deuxième Bleue

P: phosphorylase B (92 kDa)

M: marqueur de poids moléculaire

Fig. 5

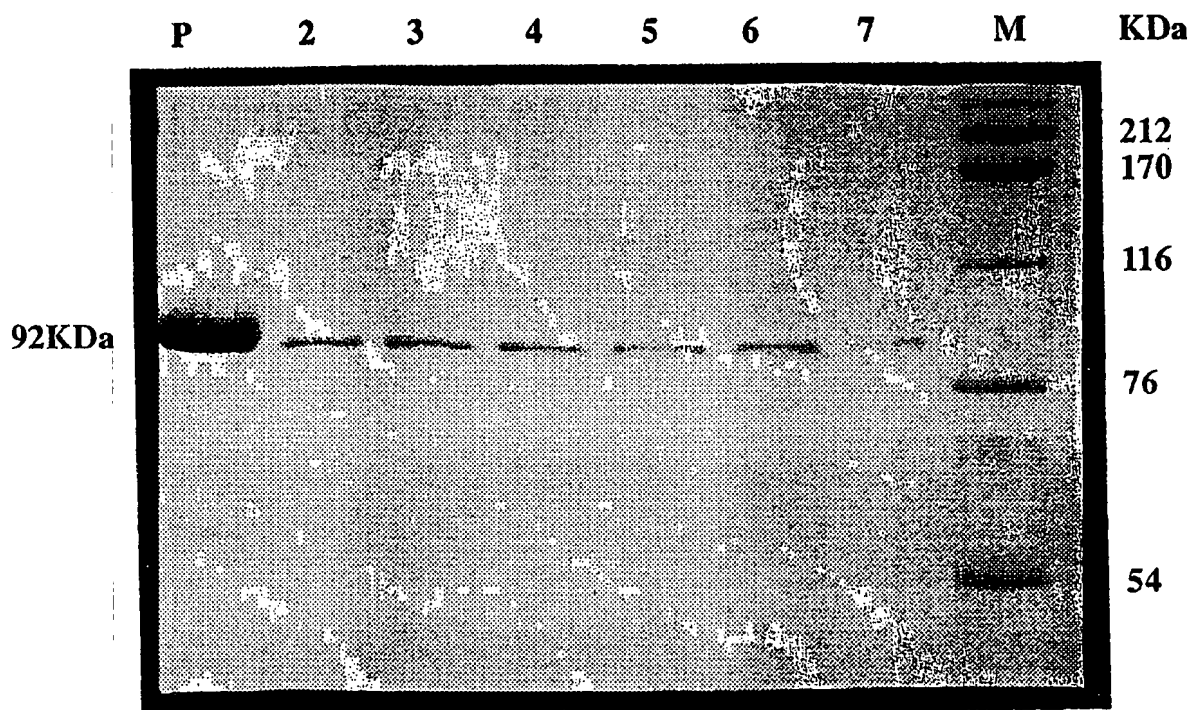


P: phosphorylase B (92 kDa)

46, 47, 48, 49: fractions après phosphocellulose

M: marqueur de poids moléculaire

Fig. 6

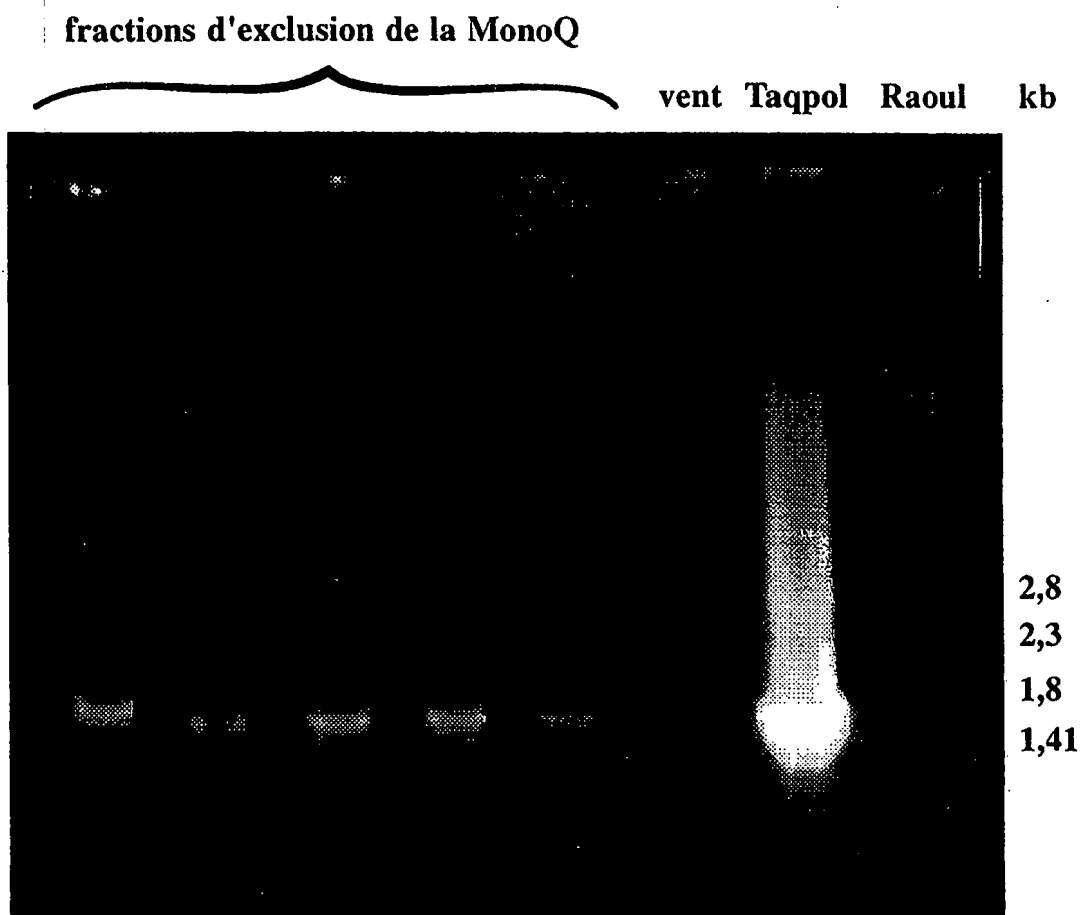


P: phosphorylase B (92 kDa)

2, 3, 4, 5, 6, 7: fractions d'exclusions sur MonoQ

M: marqueur de poids moléculaire

Fig.7



INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 97/00761

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/12 //C12N15/54

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| X | PIETROKOVSKI S: "A new intein in cyanobacteria and its significance for the spread of inteins" TRENDS IN GENETICS, vol. 12, no. 8, August 1996, page 287-288 XP004037128 see the whole document --- -/-- | 11 |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search

9 December 1997

Date of mailing of the international search report

09.01.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3018

Authorized officer

Panzica, G

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR 97/00761

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| Y | GODFROY A. ET AL.: "Thermococcus fumicolans sp. nov.: a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent in the north Fiji basin" INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, vol. 46, no. 4, October 1996, WASHINGTON US, pages 1113-1119, XP002049496 cited in the application see the whole document --- | 1-10 |
| Y | EP 0 701 000 A (NEW ENGLAND BIOLABS INC) 13 March 1996 see page 11, line 20 - page 13, line 54 --- | 1-10 |
| Y | PATENT ABSTRACTS OF JAPAN vol. 096, no. 007, 31 July 1996 & JP 08 070864 A (TOYOCO CO LTD), 19 March 1996, see abstract --- | 11 |
| Y | HODGES R. A. ET AL: "PROTEIN SPLICING REMOVES INTERVENING SEQUENCES IN AN ARCHAEA DNA POLYMERASE" NUCLEIC ACIDS RESEARCH, vol. 20, no. 23, 1992, pages 6153-6157, XP002033279 cited in the application see the whole document --- | 11 |
| A | PERLER F. B. ET AL: "INTERVENING SEQUENCES IN AN ARCHAEA DNA POLYMERASE GENE" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 89, June 1992, pages 5577-5581, XP002033280 --- | 11 |
| A | EP 0 602 899 A (NEW ENGLAND BIOLABS INC) 22 June 1994 see abstract see page 4, line 40 - page 5, line 44 ----- | 11 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 97/00761

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|----------------------------|---------------------|
| EP 0701000 A | 13-03-96 | JP 8168376 A | 02-07-96 |
| ----- | | | |
| EP 0602899 A | 22-06-94 | JP 7070200 A | 14-03-95 |
| | | CA 2110938 A | 10-06-94 |
| ----- | | | |

RAPPORT DE RECHERCHE INTERNATIONALE

De. ande internationale No
PCT/FR 97/00761

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 C12N9/12 //C12N15/54

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 6 C12N

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

| Catégorie | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
|-----------|--|-------------------------------|
| X | <p>PIETROKOVSKI S: "A new intein in cyanobacteria and its significance for the spread of inteins" TRENDS IN GENETICS, vol. 12, no. 8, août 1996, page 287-288 XP004037128 voir le document en entier ---</p> <p style="text-align: center;">-/--</p> | 11 |

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- *A* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- *E* document antérieur, mais publié à la date de dépôt international ou après cette date
- *L* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- *O* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- *P* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

T document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

X document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

Y document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

Z document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

9 décembre 1997

Date d'expédition du présent rapport de recherche internationale

09.01.98

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Panzica, G

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No

PCT/FR 97/00761

| C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS | | |
|---|---|-------------------------------|
| Catégorie | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents | no. des revendications visées |
| Y | <p>GODFROY A. ET AL.: "Thermococcus fumicolans sp. nov.: a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent in the north Fiji basin"</p> <p>INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, vol. 46, no. 4, octobre 1996, WASHINGTON US, pages 1113-1119, XP002049496</p> <p>cité dans la demande</p> <p>voir le document en entier</p> <p style="text-align: center;">---</p> | 1-10 |
| Y | <p>EP 0 701 000 A (NEW ENGLAND BIOLABS INC)</p> <p>13 mars 1996</p> <p>voir page 11, ligne 20 - page 13, ligne 54</p> <p style="text-align: center;">---</p> | 1-10 |
| Y | <p>PATENT ABSTRACTS OF JAPAN</p> <p>vol. 096, no. 007, 31 juillet 1996</p> <p>& JP 08 070864 A (TOYOBO CO LTD), 19 mars 1996,</p> <p>voir abrégé</p> <p style="text-align: center;">---</p> | 11 |
| Y | <p>HODGES R. A. ET AL: "PROTEIN SPLICING REMOVES INTERVENING SEQUENCES IN AN ARCHAEA DNA POLYMERASE"</p> <p>NUCLEIC ACIDS RESEARCH, vol. 20, no. 23, 1992, pages 6153-6157, XP002033279</p> <p>cité dans la demande</p> <p>voir le document en entier</p> <p style="text-align: center;">---</p> | 11 |
| A | <p>PERLER F. B. ET AL: "INTERVENING SEQUENCES IN AN ARCHAEA DNA POLYMERASE GENE"</p> <p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 89, juin 1992, pages 5577-5581, XP002033280</p> <p style="text-align: center;">---</p> | 11 |
| A | <p>EP 0 602 899 A (NEW ENGLAND BIOLABS INC)</p> <p>22 juin 1994</p> <p>voir abrégé</p> <p>voir page 4, ligne 40 - page 5, ligne 44</p> <p style="text-align: center;">-----</p> | 11 |

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale No

PCT/FR 97/00761

| Document brevet cité au rapport de recherche | Date de publication | Membre(s) de la famille de brevet(s) | Date de publication |
|---|------------------------|---|------------------------|
| EP 0701000 A | 13-03-96 | JP 8168376 A | 02-07-96 |
| EP 0602899 A | 22-06-94 | JP 7070200 A | 14-03-95 |
| | | CA 2110938 A | 10-06-94 |